

Cyclin-dependent kinase inhibitors and uses thereof

The present invention relates to DNA sequences encoding cyclin-dependent kinase inhibitors as well as to methods for obtaining the same. The present invention also provides vectors comprising said DNA sequences, wherein the DNA sequences are operatively linked to regulatory elements allowing expression in prokaryotic and/or eukaryotic host cells. In addition, the present invention relates to the proteins encoded by said DNA sequences, antibodies to said proteins and methods for their production. Furthermore, the present invention relates to regulatory sequences which naturally regulate the expression of the above described DNA sequences. The present invention also relates to a method for controlling or altering growth characteristics of a plant and/or a plant cell comprising introduction and/or expression of one or more cyclin-dependent kinase inhibitors functional in a plant or parts thereof and/or one or more DNA sequences encoding such proteins. Also provided by the present invention is a process for disruption plant cell division by interfering in the expression of a substrate for cyclin-dependent protein kinase using a DNA sequence according to the invention wherein said plant cell is part of a transgenic plant. The present invention further relates to diagnostic compositions comprising the aforementioned DNA sequences, proteins and antibodies. The present invention also relates to methods for the identification of compounds being capable of activating or inhibiting the cell cycle. Furthermore, the present invention relates to transgenic plant cells, plant tissue and plants containing the above-described DNA sequences and vectors as well as to the use of the aforementioned DNA sequences, vectors, proteins, antibodies, regulatory sequences and/or compounds identified by the method of the invention in plant cell and tissue culture, plant breeding and/or agriculture.

Cell division is fundamental for growth in humans, animals and plants. Prior to dividing in two daughter cells, the mother cell needs to replicate its DNA. The cell cycle is traditionally divided into 4 distinct phases:

G1 : the gap between mitosis and the onset of DNA synthesis;

S : the phase of DNA synthesis;

G2 : the gap between S and mitosis;

M : mitosis, the process of nuclear division leading up to the actual cell division.

The distinction of these 4 phases provides a convenient way of dividing the interval between successive divisions. Although they have served a useful purpose, a recent flurry of experimental results, much of it as a consequence of cancer research, has resulted in a more intricate picture of the cell cycle's "four seasons" (Nasmyth, *Science* 274, 1643-1645, 1996; Nurse, *Nature*, 344, 503-508, 1990). The underlying mechanism controlling the cell cycle control system has only recently been studied in greater detail. In all eukaryotic systems, including plants, this control mechanism is based on two key families of proteins which regulate the essential process of cell division, namely protein kinases (cyclin-dependent kinases or CDKs) and their activating associated subunits, called cyclins. The activity of these protein complexes is switched on and off at specific points of the cell cycle. Particular CDK-cyclin complexes activated at the G1/S transition trigger the start of DNA replication. Different CDK-cyclin complexes are activated at the G2/M transition and induce mitosis leading to cell division. Each of the CDK-cyclin complexes execute their regulatory role via modulating different sets of multiple target proteins. Furthermore, the large variety of developmental and environmental signals affecting cell division all converge on the regulation of CDK activity. CDKs can therefore be seen as the central engine driving cell division.

In animal systems and in yeast, knowledge about cell cycle regulations is now quite advanced. The activity of CDK-cyclin complexes is regulated at five levels: (i) transcription of the CDK and cyclin genes; (ii) association of specific CDK's with their specific cyclin partner; (iii) phosphorylation/dephosphorylation of the CDK and cyclins; (iv) interaction with other regulatory proteins such as SUC1/CKS1 homologues and cell cycle kinase inhibitors (CKI); and (v) cell cycle phase-dependent destruction of the cyclins and CKIs.

The study of cell cycle regulation in plants has lagged behind that in animals and yeast. Some basic mechanisms of cell cycle control appear to be conserved among eukaryotes, including plants. Plants were shown to also possess CDK's, cyclins and

CKI's. However plants have unique developmental features which are reflected in specific characteristics of the cell cycle control. These include for instance the absence of cell migration, the formation of organs throughout the entire lifespan from specialized regions called *meristems*, the formation of a cell wall and the capacity of non-dividing cells to re-enter the cell cycle. Another specific feature is that many plant cells, in particular those involved in storage (e.g. endosperm), are polyploid due to rounds of DNA synthesis without mitosis. This so-called endoreduplication is intimately related with cell cycle control.

Due to these fundamental differences, multiple components of the cell cycle of plants are unique compared to their yeast and animal counterparts. For example, plants contain a unique class of CDKs, such as CDC2b in *Arabidopsis*, which are both structurally and functionally different from animal and yeast CDKs. The further elucidation of cell cycle regulation in plants and its differences and similarities with other eukaryotic systems is a major research challenge. Strictly for the case of comparison, some key elements about yeast and animal systems are described below in more detail.

As already mentioned above, the control of cell cycle progression in eukaryotes is mainly exerted at two transition points: one in late G₁, before DNA synthesis, and one at the G₂/M boundary. Progression through these control points is mediated by cyclin-dependent protein kinase (CDK) complexes, which contain, in more detail, a catalytic subunit of approximately 34-kDa encoded by the *CDK* genes. Both *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* only utilize one *CDK* gene for the regulation of their cell cycle. The kinase activity of their gene products p34^{CDC2} and p34^{CDC28} in *Sch. pombe* and in *S. cerevisiae*, respectively, is dependent on regulatory proteins, called cyclins. Progression through the different cell cycle phases is achieved by the sequential association of p34^{CDC2/CDC28} with different cyclins. Although in higher eukaryotes this regulation mechanism is conserved, the situation is more complex since they have evolved to use multiple CDKs to regulate the different stages of the cell cycle. In mammals, seven CDKs have been described, defined as CDK1 to CDK7, each binding a specific subset of cyclins.

In animal systems, CDK activity is not only regulated by its association with cyclins but also involves both stimulatory and inhibitory phosphorylations. Kinase activity is

positively regulated by phosphorylation of a Thr residue located between amino acids 160-170 (depending on the CDK protein). This phosphorylation is mediated by the CDK-activating kinase (CAK) which interestingly is a CDK/cyclin complex itself. Inhibitory phosphorylations occur at the ATP-binding site (the Tyr15 residue together with Thr14 in higher eukaryotes) and are carried out by at least two protein kinases. A specific phosphatase, CDC25, dephosphorylates these residues at the G₂/M checkpoint, thus activating CDK activity and resulting in the onset of mitosis. CDK activity is furthermore negatively regulated by a family of mainly low-molecular weight proteins, called cyclin-dependent kinase inhibitors (CKIs). Kinase activity is inhibited by the tight association of these CKIs with the CDK/cyclin complexes. Cdk activity is furthermore negatively regulated by a family of mainly low-molecular weight proteins, called cyclin-dependent kinase inhibitors (CKIs). Kinase activity is inhibited by the tight association of these CKIs with the CDK/cyclin complexes. CKIs are produced during development when further cell division has to be prevented. In mammals CKIs have been shown to be involved in many different aspects of cell division and cell differentiation. First, CKI expression has been demonstrated to be induced under stress conditions such as for instance irradiation of cells or the influence of carcinogenic agents, which both potentially damage DNA. This arrest allows DNA to be repaired prior to DNA replication and mitosis. Second, inhibition of CDKs by CKIs has been demonstrated to correlate with cell differentiation and inhibition of programmed cell death. Third, the knock-out of certain members of the CKI family in mice results in an increase of body size and formation of tumors.

With respect to cell cycle regulation in plants a summary of the state of the art is given below. In *Arabidopsis*, thusfar only two CDK genes have been isolated, *CDC2aAt* and *CDC2bAt*, of which the gene products share 56% amino acid identity. Both CDKs are distinguished by several features. First, only *CDC2aAt* is able to complement yeast p34^{CDC2/CDC28} mutants. Second, *CDC2aAt* and *CDC2bAt* bear different cyclin-binding motifs (PSTAIRE and PPTALRE, respectively), suggesting they may bind distinct types of cyclins. Third, although both *CDC2aAt* and *CDC2bAt* show the same spatial expression pattern, they exhibit a different cell cycle phase-specific regulation. The *CDC2aAt* gene is expressed constitutively throughout the whole cell cycle. In contrast, *CDC2bAt* mRNA levels oscillate, being most abundant

during the S and G₂ phases. In addition, multiple cyclins have been isolated from *Arabidopsis*. The majority displays the strongest sequence similarity with the animal A- or B-type class of cyclins, but also D-type cyclins have been identified. Although the classification of *Arabidopsis* cyclins is mainly based upon sequence similarity, limited data suggests that this organization corresponds with differential functions of each cyclin class. Recently, a CDK inhibitor has been identified in *Arabidopsis thaliana* (ICK1) that shares some limited similarity with the mammalian p27^{kip1} kinase inhibitor (Wang, Nature 386 (1997), 451-452). This CDK inhibitor was predominantly identified when screening a library with a yeast two-hybrid "bait" construct harboring *Arabidopsis thaliana* CDC2aAt cDNA suggesting that only one class of CDK inhibitors is present in plants. However, the function and expression of CDK inhibitors in plants still needs to be determined.

In order to manage problems related to plant growth, plant architecture and/or plant diseases, it is believed to be of utmost importance to identify, isolate plant and characterize genes and gene products involved in the regulation of the plant cell division, and more particularly coding for and interacting with CDK's and/or their interacting proteins, responsible for the control of the cell cycle and the completion of the S and M phase of the cell cycle. If such novel genes and/or proteins have been isolated and analyzed, the growth of the plant as a whole can be influenced. Also, the growth of specific tissues or organs and thus the architecture of the plant can be modified.

Thus, the technical problem underlying the present invention is to provide means and methods for modulating cell cycle proteins that are particular useful in agriculture and plant cell and tissue culture.

The solution to the technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the invention relates to a DNA sequence encoding a cyclin-dependent kinase inhibitor or encoding an immunologically active and/or functional fragment of such a protein, selected from the group consisting of:

- (a) DNA sequences comprising a nucleotide sequence encoding a protein comprising the amino acid sequence as given in SEQ ID NO: 2, 4 or 6;
- (b) DNA sequences comprising a nucleotide sequence as given in SEQ ID NO: 1, 3 or 5;
- (c) DNA sequences comprising the nucleotide sequence encoding a protein comprising the amino acid sequence from amino acid position 75 to 209 of SEQ ID NO: 2 or from amino acid position 11 to 216 of SEQ ID NO: 4 or comprising the nucleotide sequence from nucleotide position 305 to 932 of SEQ ID NO: 1;
- (d) DNA sequences hybridizing with the complementary strand of a DNA sequence as defined in any one of (a) to (c);
- (e) DNA sequences encoding an amino acid sequence which is at least 30% identical to the amino acid sequence encoded by the DNA sequence of any one of (a) to (c);
- (f) DNA sequences, the nucleotide sequence of which is degenerated as a result of the genetic code to a nucleotide sequence of a DNA sequence as defined in any one of (a) to (e); and
- (g) DNA sequences encoding a fragment of a protein encoded by a DNA sequence of any one of (a) to (f).

The term "cyclin-dependent kinase inhibitor" also designated CDK inhibitor, CKI or CDKI as denoted herein means a protein which inhibits CDK/cyclin activity and is produced during development when further cell division has to be prevented. A CDK inhibitor of the invention is capable of inhibiting or suppressing the kinase activity of protein kinases, in particular of cyclin-dependent kinases. The capability of a inhibiting or suppressing protein kinase activity can be determined according to methods well known in the art; see, e.g., Wang, *supra* and the appended examples. The term "cell cycle" means the cyclic biochemical and structural events associated with growth of cells, and in particular with the regulation of the replication of DNA

and mitosis. The cycle is divided into periods called: G_0 , Gap_1 (G_1), DNA synthesis (S), Gap_2 (G_2), and mitosis (M).

The terms "gene(s)", "polynucleotide", "nucleic acid sequence", "nucleotide sequence", "DNA sequence" or "nucleic acid molecule(s)" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, and RNA. It also includes known types of modifications, for example, methylation, "caps" substitution of one or more of the naturally occurring nucleotides with an analog. Preferably, the DNA sequence of the invention comprises a coding sequence encoding the above defined cell cycle interacting protein.

A "coding sequence" is a nucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, recombinant nucleotide sequences or genomic DNA, while introns may be present as well under certain circumstances.

In accordance with the present invention new plant gene products with a putative CDK inhibitory function were screened by using the two-hybrid system (Fields, Nature 340 (1989), 245-246). For this purpose the CDC2aAt protein was exploited as bait. Previous attempts using the identical bait and a cDNA library constructed with RNA from 3-week-old *Arabidopsis thaliana* vegetative tissues were unsuccessful (De Veylder, Febs Lett. 412 (1997), 446-452; De Veylder, J. Exp. Bot. 48 (1997), 2113-2114). A new attempt was undertaken using a newly constructed library made from a RNA mixture of *Arabidopsis thaliana* cell suspensions harvested at various growing stages: early exponential, exponential, early stationary and stationary phase. This library has the advantage above the previous one to include mainly genes expressed in cells at the onset of cell division, actively dividing cells, cells redrawing from the cell cycle, and non-cycling cells. Surprisingly, using this specific library several positive clones were identified encoding proteins with a

putative CDK inhibitory function. These clones were designated LDV39, LDV66, and LDV159.

A homology search in databases revealed that the last 23 amino-acids showed significant homology to the human CKIs p21^{cip1} and p27^{kip1}. The LDV39 gene was 622 bp long, consisting of 423 bp coding region and 199 bp 3' UTR (excluding the poly-A tail). The LDV66 gene was 611 bp long, consisting of 379 bp coding region and 232 bp 3' UTR (excluding the poly-A tail). Since the *LDV39* and *LDV66* clones encode partial proteins, lacking their amino-terminal part, a flower cDNA library obtained from the ABRC stock centre (library stock number CD4-6) was screened. The positive clones were denominated *FL39* and *FL66*, corresponding to longer clones of *LDV39* and *LDV66*, respectively.

The *FL39* clone is 932 bp (SEQ ID NO: 1) long and contains an ORF encoding a protein of 209 amino acids (SEQ ID NO: 2) with a calculated molecular mass of 24 kDa. In its 3' UTR a poly-adenylation signal can be recognised. The amino-terminal part of the *FL39* protein contains a repeated motif of 11 amino acids (VRRRD/ExxxVEE). This motif is not found in any other protein in the databanks and its significance is unknown. The *FL39* protein also contains a putative nuclear localization signal (amino acids 23-26) and a PEST-rich region (amino acids 71-98; PESTFIND score +15.5). These sequences, rich in proline, glutamic acid, serine and proline, are characteristically present in unstable proteins (Rogers et al., 1986, Science 234, 364-368).

The *FL66* sequence does not contain an in frame stopcodon, and may therefore not be full length. The *FL66* clone is 875 bp long (SEQ ID NO: 3) and bears an ORF of 216 amino acids (SEQ ID NO: 4), encoding a protein of 24 kD. No nuclear localization signal or PEST domains are present. Furthermore, a CDK inhibitor named ALFCDKI from alfalfa has been identified in accordance with the present invention using a two-hybrid screening assay. This gene comprises 1202 nucleotides (SEQ ID NO:5) with a coding region from nucleotide position 94 to 760 encoding a protein of 224 amino acids (SEQ ID NO:6). The LDV159 clone was identical to ICK1 (GenBank accession number U94772 as published by Wang, Nature 386 (1997), p451-452). Surprisingly, the three other clones were novel and encoded proteins only distantly related to ICK1 (Table 1).

Table 1: Sequence similarity and identity between the different plant cyclin-dependent kinase inhibitors. CrCKI is the *Chenopodium rubrum* CKI (accession number AJ00217).

	FL39	FL66	FL67	ICK1	CrCKI	ALFCDKI
FL39		27.805	33.333	32.292	34.392	N.S.
FL66	<i>21.463</i>		39.545	37.017	34.574	34.389
FL67	<i>20.000</i>	<i>30.909</i>		30.220	N.S.	N.S.
ICK	<i>23.958</i>	<i>30.939</i>	<i>22.527</i>		32.105	25.131
CrCKI	<i>24.868</i>	<i>28.723</i>	<i>N.S.</i>	<i>27.368</i>		26.667
MsCKI	<i>N.S.</i>	<i>28.054</i>	<i>N.S.</i>	<i>21.990</i>	<i>20.513</i>	

The percentage similarity (bold) and identity (italic) between the different Cyclin-dependent kinase inhibitors was determined using the GAP program (from the GCG 9.1 package) using the parameters Gap weight = 12 and Length weight = 4. N.S.: Not Significant.

Furthermore, the genomic organisation of the *FL39*, *FL66* and *ICK1* clones was tested by DNA gel blot analysis. The results of the experiments suggest the presence of an additional *FL66* related gene and, therefore, it can be concluded that there are at least four different CKI proteins present in *A. thaliana*. From the foregoing it is evident that more than one CDK inhibitor in plants exist and therefore different functions during plant development and/or expression patterns can be assumed. Further studies that have been performed in accordance with the present invention revealed that the CDK inhibitors are expressed at different time points during the cultivation of the plant cell culture; see Example 8. Moreover, it could be demonstrated in accordance with the present invention that the CDK inhibitor *FL66* is regulated by NaCl; see Example 9. The inhibitory function of the CDK inhibitor of the invention is exemplified with *FL66*; see Example 6. In addition, in situ hybridization using antisense probes derived from cDNAs from LDV39, LDV66 and LDV159 demonstrated that each of these CDK inhibitors exhibit distinct expression patterns; see Example 13. Thus, the findings of the present invention establishes that in plants several CDK inhibitors exist which due to their differential expression

pattern may have different functions during the development of the plant. It can be expected that similar gene families encoding CDK inhibitors are present in other plant species than *Arabidopsis* and alfalfa as well. These cyclin-dependent inhibitors are also within the scope of the present invention.

Accordingly, the present invention also relates to nucleic acid molecules hybridizing with the above-described nucleic acid molecules and differ in one or more positions in comparison with these as long as they encode a cyclin-dependent kinase inhibitor. By "hybridizing" it is meant that such nucleic acid molecules hybridize under conventional hybridization conditions, preferably under stringent conditions such as described by, e.g., Sambrook (Molecular Cloning; A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). Preferably, the hybridization conditions used in the examples are employed. Cyclin-dependent kinase inhibitor derived from other organisms such as mammals, in particular humans, may be encoded by other DNA sequences which hybridize to the sequences for plant cyclin-dependent kinase inhibitor under relaxed hybridization conditions and which code on expression for peptides having the ability to interact with cell cycle proteins. Examples of such non-stringent hybridization conditions are 4XSSC at 50 °C or hybridization with 30-40 % formamide at 42 °C. Such molecules comprise those which are fragments, analogues or derivatives of the cell cycle interacting protein of the invention and differ, for example, by way of amino acid and/or nucleotide deletion(s), insertion(s), substitution(s), addition(s) and/or recombination(s) or any other modification(s) known in the art either alone or in combination from the above-described amino acid sequences or their underlying nucleotide sequence(s). Methods for introducing such modifications in the nucleic acid molecules according to the invention are well-known to the person skilled in the art. The invention also relates to nucleic acid molecules the sequence of which differs from the nucleotide sequence of any of the above-described nucleic acid molecules due to the degeneracy of the genetic code. All such fragments, analogues and derivatives of the protein of the invention are included within the scope of the present invention, as long as the essential characteristic immunological and/or biological properties as defined above remain unaffected in kind, that is the

novel nucleic acid molecules of the invention include all nucleotide sequences encoding proteins or peptides which have at least a part of the primary structural conformation for one or more epitopes capable of reacting with antibodies to cyclin-dependent kinase inhibitor which are encodable by a nucleic acid molecule as set forth above and which have comparable or identical characteristics in terms of inhibiting cyclin dependent kinases, in particular plant cyclin dependent kinases. Part of the invention are therefore also nucleic acid molecules encoding a polypeptide comprising at least a functional part of cyclin-dependent kinase inhibitor encoded by a nucleic acid sequence comprised in a nucleic acid molecule according to the invention. An example for this is that the polypeptide or a fragment thereof according to the invention is embedded in another amino acid sequence.

As is demonstrated in the appended examples a two-hybrid screening assay has been developed in accordance with the present invention suitable for identifying cyclin-dependent kinase inhibitor. Thus, in another aspect the present invention relates to a method for identifying and obtaining cyclin-dependent kinase inhibitors comprising a two-hybrid screening assay wherein CDC2a as a bait and a cDNA library of cell suspension as prey are used. Preferably, said CDC2a is CDC2aAt. However, CDC2a from other organisms such as other plants but also mammals may be employed as well.

The nucleic acid molecules encoding proteins or peptides identified to interact with the CDC2a in the above mentioned assay can be easily obtained and sequenced by methods known in the art; see also the appended examples. Therefore, the present invention also relates to a DNA sequence encoding a cyclin-dependent kinase inhibitor obtainable by the method of the invention. Preferably, the amino acid sequence of said protein obtainable by the method of the invention has an identity to the amino acid sequence of any one of SEQ ID NOS: 2, 4 or 6 of at least 30 %, more preferably 40 to 60 % and most preferably 70 % to 90 %.

In a preferred embodiment the nucleic acid molecules according to the invention are RNA or DNA molecules, preferably cDNA, genomic DNA or synthetically

synthesized DNA or RNA molecules. Preferably, the nucleic acid molecule of the invention is derived from a plant, preferably from *Arabidopsis thaliana*. As discussed above, a cyclin-dependent kinase inhibitor could also be identified in *Medicago sativa* (Alfalfa). Corresponding proteins displaying similar properties should, therefore, be present in other plants as well. Nucleic acid molecules of the invention can be obtained, e.g., by hybridization of the above-described nucleic acid molecules with a (sample of) nucleic acid molecule(s) of any source. Nucleic acid molecules hybridizing with the above-described nucleic acid molecules can in general be derived from any organism, preferably plant possessing such molecules, preferably from monocotyledonous or dicotyledonous plants, in particular from any organism, preferably plants of interest in agriculture, horticulture or wood culture, such as crop plants, namely those of the family Poaceae, any starch producing plants, such as potato, maniok, leguminous plants, oil producing plants, such as oilseed rape, linenseed, etc., plants using polypeptide as storage substances, such as soybean, plants using sucrose as storage substance, such as sugar beet or sugar cane, trees, ornamental plants etc. Preferably, the nucleic acid molecules according to the invention are derived from *Arabidopsis thaliana*. Nucleic acid molecules hybridizing to the above-described nucleic acid molecules can be isolated, e.g., from libraries, such as cDNA or genomic libraries by techniques well known in the art. For example, hybridizing nucleic acid molecules can be identified and isolated by using the above-described nucleic acid molecules or fragments thereof or complements thereof as probes to screen libraries by hybridizing with said molecules according to standard techniques. Possible is also the isolation of such nucleic acid molecules by applying the polymerase chain reaction (PCR) using as primers oligonucleotides derived from the above-described nucleic acid molecules.

Nucleic acid molecules which hybridize with any of the aforementioned nucleic acid molecules also include fragments, derivatives and allelic variants of the above-described nucleic acid molecules that encode a cyclin-dependent kinase inhibitor or an immunologically or functional fragment thereof. Fragments are understood to be parts of nucleic acid molecules long enough to encode the described protein or a

functional or immunologically active fragment thereof as defined above. Preferably, the functional fragment contains a motif of 11 amino acids (VRRRD/ExxxVEE; SEQ ID NO: 33) present in the amino terminal part of the FL39 protein. This motif is not found in any other protein in the databanks and its significance is unknown. Furthermore, the fragment may contain the putative nuclear localization signal (amino acids 23-26 of SEQ ID NO: 2) and/or the PEST-rich region (amino acids 71-98 of SEQ ID NO: 2; see also Example 3).

The term "derivative" means in this context that the nucleotide sequence of these nucleic acid molecules differs from the sequences of the above-described nucleic acid molecules in one or more nucleotide positions and are highly homologous to said nucleic acid molecules. Homology is understood to refer to a sequence identity of at least 30 %, particularly an identity of at least 60 %, preferably more than 80 % and still more preferably more than 90 %. The term "substantially homologous" refers to a subject, for instance a nucleic acid, which is at least 50% identical in sequence to the reference when the entire ORF (open reading frame) is compared, where the sequence identity is preferably at least 70%, more preferably at least 80%, still more preferably at least 85%, especially more than about 90%, most preferably 95% or greater, particularly 98% or greater. The deviations from the sequences of the nucleic acid molecules described above can, for example, be the result of nucleotide substitution(s), deletion(s), addition(s), insertion(s) and/or recombination(s); see supra.

Homology further means that the respective nucleic acid molecules or encoded proteins are functionally and/or structurally equivalent. The nucleic acid molecules that are homologous to the nucleic acid molecules described above and that are derivatives of said nucleic acid molecules are, for example, variations of said nucleic acid molecules which represent modifications having the same biological function, in particular encoding proteins with the same or substantially the same biological function. They may be naturally occurring variations, such as sequences from other plant varieties or species, or mutations. These mutations may occur naturally or may be obtained by mutagenesis techniques. The allelic variations may

be naturally occurring allelic variants as well as synthetically produced or genetically engineered variants; see supra.

The proteins encoded by the various derivatives and variants of the above-described nucleic acid molecules share specific common characteristics, such as biological activity, molecular weight, immunological reactivity, conformation, etc., as well as physical properties, such as electrophoretic mobility, chromatographic behavior, sedimentation coefficients, pH optimum, temperature optimum, stability, solubility, spectroscopic properties, etc.

Examples of the different possible applications of the nucleic acid molecules according to the invention as well as molecules derived from them will be described in detail in the following.

Hence, in a further embodiment, the invention relates to nucleic acid molecules of at least 15 nucleotides in length hybridizing specifically with a nucleic acid molecule as described above or with a complementary strand thereof. Specific hybridization occurs preferably under stringent conditions and implies no or very little cross-hybridization with nucleotide sequences encoding no or substantially different proteins. Such nucleic acid molecules may be used as probes and/or for the control of gene expression. Nucleic acid probe technology is well known to those skilled in the art who will readily appreciate that such probes may vary in length. Preferred are nucleic acid probes of 16 to 35 nucleotides in length. Of course, it may also be appropriate to use nucleic acids of up to 100 and more nucleotides in length. The nucleic acid probes of the invention are useful for various applications. On the one hand, they may be used as PCR primers for amplification of nucleic acid sequences according to the invention. The design and use of said primers is known by the person skilled in the art. Preferably such amplification primers comprise a contiguous sequence of at least 6 nucleotides, in particular 13 nucleotides, preferably 15 to 25 nucleotides or more, identical or complementary to the nucleotide sequence depicted in SEQ ID NO: 1, 3 or 5 or to a nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 2, 4 or 6. Another application is

the use as a hybridization probe to identify nucleic acid molecules hybridizing with a nucleic acid molecule of the invention by homology screening of genomic DNA or cDNA libraries. Nucleic acid molecules according to this preferred embodiment of the invention which are complementary to a nucleic acid molecule as described above may also be used for repression of expression of a CKI encoding gene, for example due to an antisense or triple helix effect or for the construction of appropriate ribozymes (see, e.g., EP-A1 0 291 533, EP-A1 0 321 201, EP-A2 0 360 257) which specifically cleave the (pre)-mRNA of a gene comprising a nucleic acid molecule of the invention or part thereof. Selection of appropriate target sites and corresponding ribozymes can be done as described, for example, in Steinecke, Ribozymes, Methods in Cell Biology 50, Galbraith et al. eds Academic Press, Inc. (1995), 449-460. Furthermore, the person skilled in the art is well aware that it is also possible to label such a nucleic acid probe with an appropriate marker for specific applications, such as for the detection of the presence of a nucleic acid molecule of the invention in a sample derived from an organism, in particular plants.

The above described nucleic acid molecules may either be DNA or RNA or a hybrid thereof. Furthermore, said nucleic acid molecule may contain, for example, thioester bonds and/or nucleotide analogues, commonly used in oligonucleotide anti-sense approaches. Said modifications may be useful for the stabilization of the nucleic acid molecule against endo- and/or exonucleases in the cell. Said nucleic acid molecules may be transcribed by an appropriate vector containing a chimeric gene which allows for the transcription of said nucleic acid molecule in the cell.

Furthermore, the so-called "peptide nucleic acid" (PNA) technique can be used for the detection or inhibition of the expression of a nucleic acid molecule of the invention. For example, the binding of PNAs to complementary as well as various single stranded RNA and DNA nucleic acid molecules can be systematically investigated using thermal denaturation and BIAcore surface-interaction techniques (Jensen, Biochemistry 36 (1997), 5072-5077). Furthermore, the nucleic acid molecules described above as well as PNAs derived therefrom can be used for detecting point mutations by hybridization with nucleic acids obtained from a sample

with an affinity sensor, such as BIAcore; see Gotoh, *Rinsho Byori* 45 (1997), 224-228. Hybridization based DNA screening on peptide nucleic acids (PNA) oligomer arrays are described in the prior art, for example in Weiler, *Nucleic Acids Research* 25 (1997), 2792-2799. The synthesis of PNAs can be performed according to methods known in the art, for example, as described in Koch, *J. Pept. Res.* 49 (1997), 80-88; Finn, *Nucleic Acids Research* 24 (1996), 3357-3363. Further possible applications of such PNAs, for example as restriction enzymes or as templates for the synthesis of nucleic acid oligonucleotides are known to the person skilled in the art and are, for example, described in Veselkov, *Nature* 379 (1996), 214 and Bohler, *Nature* 376 (1995), 578-581.

The present invention also relates to vectors, particularly plasmids, cosmids, viruses, bacteriophages and other vectors used conventionally in genetic engineering that contain a nucleic acid molecule according to the invention. Methods which are well known to those skilled in the art can be used to construct various plasmids and vectors; see, for example, the techniques described in Sambrook, *Molecular Cloning A Laboratory Manual*, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the nucleic acid molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

In a preferred embodiment the nucleic acid molecule present in the vector is linked to (a) control sequence(s) which allow the expression of the nucleic acid molecule in prokaryotic and/or eukaryotic cells.

The term "control sequence" refers to regulatory DNA sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoter, ribosomal binding site, and terminators. In eukaryotes generally control sequences include promoters, terminators and, in some instances, enhancers, transactivators or transcription factors. The term "control sequence" is intended to include, at a minimum, all

components the presence of which are necessary for expression, and may also include additional advantageous components.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is used.

Thus, the vector of the invention is preferably an expression vector. An "expression vector" is a construct that can be used to transform a selected host cell and provides for expression of a coding sequence in the selected host. Expression vectors can for instance be cloning vectors, binary vectors or integrating vectors. Expression comprises transcription of the nucleic acid molecule preferably into a translatable mRNA. Regulatory elements ensuring expression in prokaryotic and/or eukaryotic cells are well known to those skilled in the art. In the case of eukaryotic cells they comprise normally promoters ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript, for example, those of the 35S RNA from Cauliflower Mosaic Virus (CaMV). Other promoters commonly used are the polyubiquitin promoter, and the actin promoter for ubiquitous expression. The termination signals usually employed are from the Nopaline Synthase promoter or from the CAMV 35S promoter. A plant translational enhancer often used is the CAMV omega sequences, the inclusion of an intron (Intron-1 from the Shrunken gene of maize, for example) has been shown to increase expression levels by up to 100-fold. (Mait, Transgenic Research 6 (1997), 143-156; Ni, Plant Journal 7 (1995), 661-676). Additional regulatory elements may include transcriptional as well as translational enhancers. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the P_L , *lac*, *trp* or *tac* promoter in *E. coli*, and examples of regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. In this context, suitable expression vectors are known in the art such as Okayama-Berg

cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogen), pSPORT1 (GIBCO BRL). Advantageously, the above-described vectors of the invention comprises a selectable and/or scorable marker. Selectable marker genes useful for the selection of transformed plant cells, callus, plant tissue and plants are well known to those skilled in the art and comprise, for example, antimetabolite resistance as the basis of selection for dhfr, which confers resistance to methotrexate (Reiss, *Plant Physiol. (Life Sci. Adv.)* 13 (1994), 143-149); npt, which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (Herrera-Estrella, *EMBO J.* 2 (1983), 987-995) and hygromycin (Marsh, *Gene* 32 (1984), 481-485). Additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histidinol in place of histidine (Hartman, *Proc. Natl. Acad. Sci. USA* 85 (1988), 8047); mannose-6-phosphate isomerase which allows cells to utilize mannose (WO 94/20627) and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, 1987, In: *Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory ed.) or deaminase from *Aspergillus terreus* which confers resistance to Blasticidin S (Tamura, *Biosci. Biotechnol. Biochem.* 59 (1995), 2336-2338).

Useful scorable marker are also known to those skilled in the art and are commercially available. Advantageously, said marker is a gene encoding luciferase (Giacomin, *Pl. Sci.* 116 (1996), 59-72; Scikantha, *J. Bact.* 178 (1996), 121), green fluorescent protein (Gerdes, *FEBS Lett.* 389 (1996), 44-47) or β -glucuronidase (Jefferson, *EMBO J.* 6 (1987), 3901-3907). This embodiment is particularly useful for simple and rapid screening of cells, tissues and organisms containing a vector of the invention.

The present invention furthermore relates to host cells comprising a vector as described above or a nucleic acid molecule according to the invention wherein the nucleic acid molecule is foreign to the host cell.

By "foreign" it is meant that the nucleic acid molecule is either heterologous with respect to the host cell, this means derived from a cell or organism with a different genomic background, or is homologous with respect to the host cell but located in a different genomic environment than the naturally occurring counterpart of said nucleic acid molecule. This means that, if the nucleic acid molecule is homologous with respect to the host cell, it is not located in its natural location in the genome of said host cell, in particular it is surrounded by different genes. In this case the nucleic acid molecule may be either under the control of its own promoter or under the control of a heterologous promoter. The vector or nucleic acid molecule according to the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained in some form extrachromosomally. In this respect, it is also to be understood that the nucleic acid molecule of the invention can be used to restore or create a mutant gene via homologous recombination (Paszkowski (ed.), *Homologous Recombination and Gene Silencing in Plants*. Kluwer Academic Publishers (1994)).

The host cell can be any prokaryotic or eukaryotic cell, such as bacterial, insect, fungal, plant or animal cells. Preferred fungal cells are, for example, those of the genus *Saccharomyces*, in particular those of the species *S. cerevisiae*.

Another subject of the invention is a method for the preparation of a cyclin-dependent kinase inhibitor which comprises the cultivation of host cells according to the invention which, due to the presence of a vector or a nucleic acid molecule according to the invention, are able to express such a protein, under conditions which allow expression of the protein and recovering of the so-produced protein from the culture.

The term "expression" means the production of a protein or nucleotide sequence in the cell. However, said term also includes expression of the protein in a cell-free system. It includes transcription into an RNA product, post-transcriptional modification and/or translation to a protein product or polypeptide from a DNA encoding that product, as well as possible post-translational modifications. Depending on the specific constructs and conditions used, the protein may be

recovered from the cells, from the culture medium or from both. For the person skilled in the art it is well known that it is not only possible to express a native protein but also to express the protein as fusion polypeptides or to add signal sequences directing the protein to specific compartments of the host cell, e.g., ensuring secretion of the peptide into the culture medium, etc. Furthermore, such a protein and fragments thereof can be chemically synthesized and/or modified according to standard methods described, for example hereinbelow.

The terms "protein" and "polypeptide" used in this application are interchangeable. "Polypeptide" refers to a polymer of amino acids (amino acid sequence) and does not refer to a specific length of the molecule. Thus peptides and oligopeptides are included within the definition of polypeptide. This term does also refer to or include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The present invention furthermore relates to CKIs encoded by the nucleic acid molecules according to the invention or produced or obtained by the above-described methods, and to functional and/or immunologically active fragments of such cyclin-dependent kinase inhibitor. The proteins and polypeptides of the present invention are not necessarily translated from a designated nucleic acid sequence; the polypeptides may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from a suitable viral system. The polypeptides may include one or more analogs of amino acids, phosphorylated amino acids or unnatural amino acids. Methods of inserting analogs of amino acids into a sequence are known in the art. The polypeptides may also include one or more labels, which are known to those skilled in the art. In this context, it is also understood that the proteins according to the invention may be further modified by conventional methods known in the art. By providing the proteins according to the present invention it is also possible to

determine fragments which retain biological activity, for example, the mature, processed form. This allows the construction of chimeric proteins and peptides comprising an amino sequence derived from the protein of the invention, which is crucial for its binding activity and other functional amino acid sequences, e.g. GUS marker gene (Jefferson, EMBO J. 6 (1987), 3901-3907). The other functional amino acid sequences may be either physically linked by, e.g., chemical means to the proteins of the invention or may be fused by recombinant DNA techniques well known in the art.

The term "fragment of a sequence" or "part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least a comparable function and/or activity of the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or function(s) of the original sequence. Typically, the truncated amino acid sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids. Preferably, the polypeptides according to the invention comprising the amino acid sequence as defined above and/or a fragment thereof have a molecular weight of approximately 15-20 kDa.

Furthermore, folding simulations and computer redesign of structural motifs of the protein of the invention can be performed using appropriate computer programs (Olszewski, Proteins 25 (1996), 286-299; Hoffman, Comput. Appl. Biosci. 11 (1995), 675-679). Computer modeling of protein folding can be used for the conformational and energetic analysis of detailed peptide and protein models (Monge, J. Mol. Biol. 247 (1995), 995-1012; Renouf, Adv. Exp. Med. Biol. 376 (1995), 37-45). In particular, the appropriate programs can be used for the identification of interactive sites of the CKI and cyclin dependent kinases, its ligand or other interacting proteins by computer assisted searches for complementary peptide sequences (Fassina, Immunomethods

5 (1994), 114-120). Further appropriate computer systems for the design of protein and peptides are described in the prior art, for example in Berry, *Biochem. Soc. Trans.* 22 (1994), 1033-1036; Wodak, *Ann. N. Y. Acad. Sci.* 501 (1987), 1-13; Pabo, *Biochemistry* 25 (1986), 5987-5991. The results obtained from the above-described computer analysis can be used for, e.g., the preparation of peptidomimetics of the protein of the invention or fragments thereof. Such pseudopeptide analogues of the natural amino acid sequence of the protein may very efficiently mimic the parent protein (Benkirane, *J. Biol. Chem.* 271 (1996), 33218-33224). For example, incorporation of easily available achiral Ω -amino acid residues into a protein of the invention or a fragment thereof results in the substitution of amide bonds by polymethylene units of an aliphatic chain, thereby providing a convenient strategy for constructing a peptidomimetic (Banerjee, *Biopolymers* 39 (1996), 769-777). Superactive peptidomimetic analogues of small peptide hormones in other systems are described in the prior art (Zhang, *Biochem. Biophys. Res. Commun.* 224 (1996), 327-331). Appropriate peptidomimetics of the protein of the present invention can also be identified by the synthesis of peptidomimetic combinatorial libraries through successive amide alkylation and testing the resulting compounds, e.g., for their binding, kinase inhibitory and/or immunological properties. Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh, *Methods in Enzymology* 267 (1996), 220-234 and Dorner, *Bioorg. Med. Chem.* 4 (1996), 709-715.

Furthermore, a three-dimensional and/or crystallographic structure of the protein of the invention can be used for the design of peptidomimetic inhibitors of the biological activity of the protein of the invention (Rose, *Biochemistry* 35 (1996), 12933-12944; Rutenber, *Bioorg. Med. Chem.* 4 (1996), 1545-1558).

Furthermore, the present invention relates to antibodies specifically recognizing a cyclin-dependent kinase inhibitor according to the invention or parts, i.e. specific fragments or epitopes, of such a protein. The antibodies of the invention can be used to identify and isolate other cyclin-dependent kinase inhibitors and genes in any organism, preferably plants. These antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such

as Fab, Fv or scFv fragments etc. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, *Nature* 256 (1975), 495, and Galfré, *Meth. Enzymol.* 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. Furthermore, antibodies or fragments thereof to the aforementioned peptides can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of proteins according to the invention as well as for the monitoring of the synthesis of such proteins, for example, in recombinant organisms, and for the identification of compounds interacting with the protein according to the invention. For example, surface plasmon resonance as employed in the BIAcore system can be used to increase the efficiency of phage antibodies selections, yielding a high increment of affinity from a single library of phage antibodies which bind to an epitope of the protein of the invention (Schier, *Human Antibodies Hybridomas* 7 (1996), 97-105; Malmberg, *J. Immunol. Methods* 183 (1995), 7-13). In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

Plant cell division can conceptually be influenced in three ways : (i) inhibiting or arresting cell division, (ii) maintaining, facilitating or stimulating cell division or (iii) uncoupling DNA synthesis from mitosis and cytokinesis. Modulation of the expression of a polypeptide encoded by a nucleotide sequence according to the invention has surprisingly an advantageous influence on plant cell division characteristics, in particular on the disruption of the expression levels of genes or the biological activity of the proteins involved in G1/S and/or G2/M transition and as a result thereof on the total make-up of the plant concerned or parts thereof. An example is that DNA synthesis or progression of DNA replication will be negatively influenced by inactivating or inhibiting cyclin-dependent protein kinase complexes.

The term "cyclin-dependent protein kinase complex" means the complex formed when a, preferably functional, cyclin associates with a, preferably, functional cyclin

dependent kinase. Such complexes may be active in phosphorylating proteins and may or may not contain additional protein species.

The activity of a CDK in a plant cell is influenced by manipulation of the gene according to the invention. To analyse the industrial applicabilities of the invention, transformed plants can be made overproducing the nucleotide sequence according to the invention. Such an overexpression of the new gene(s), proteins or inactivated variants thereof will either positively or negatively have an effect on cell division. Methods to modify the expression levels and/or the activity are known to persons skilled in the art and include for instance overexpression, co-suppression, the use of ribozymes, sense and anti-sense strategies, gene silencing approaches. "Sense strand" refers to the strand of a double-stranded DNA molecule that is homologous to a mRNA transcript thereof. The "anti-sense strand" contains an inverted sequence which is complementary to that of the "sense strand".

Hence, the nucleic acid molecules according to the invention are in particular useful for the genetic manipulation of plant cells in order to modify the characteristics of plants and to obtain plants with modified, preferably with improved or useful phenotypes. Similarly, the invention can also be used to modulate the cell division and the growth of cells, preferentially plant cells, in *in vitro* cultures.

Thus, the present invention provides for a method for the production of transgenic plants, plant cells or plant tissue comprising the introduction of a nucleic acid molecule or vector of the invention into the genome of said plant, plant cell or plant tissue.

For the expression of the nucleic acid molecules according to the invention in sense or antisense orientation in plant cells, the molecules are placed under the control of regulatory elements which ensure the expression in plant cells. These regulatory elements may be heterologous or homologous with respect to the nucleic acid molecule to be expressed as well with respect to the plant species to be transformed. In general, such regulatory elements comprise a promoter active in plant cells. To obtain expression in all tissues of a transgenic plant, preferably constitutive promoters are used, such as the 35 S promoter of CaMV (Odell, Nature 313 (1985), 810-812) or

promoters of the polyubiquitin genes of maize (Christensen, *Plant Mol. Biol.* 18 (1982), 675-689). Furthermore, the expression of the nucleic acid molecules of the invention can be controlled by, e.g., introduction of high constitutive, tissue specific, cell type specific or inducible promoters adjacent to said nucleotide sequence or fragment thereof, multiple gene repeats and other similar techniques. For instance transgenic plants can thus be obtained which can not form feeding cells upon nematode infection of the roots. It is also feasible to generate transgenic plants which are resistant to certain viral infections such as a gemini viral infection. In order to achieve expression in specific tissues of a transgenic plant it is possible to use tissue specific promoters (see, e.g., Stockhaus, *EMBO J.* 8 (1989), 2245-2251). Known are also promoters which are specifically active in tubers of potatoes or in seeds of different plants species, such as maize, Vicia, wheat, barley etc. Inducible promoters may be used in order to be able to exactly control expression. An example for inducible promoters are the promoters of genes encoding heat shock proteins. Also microspore-specific regulatory elements and their uses have been described (WO96/16182). Furthermore, the chemically inducible Test-system may be employed (Gatz, *Mol. Gen. Genet.* 227 (1991); 229-237). Further suitable promoters are known to the person skilled in the art and are described, e.g., in Ward (*Plant Mol. Biol.* 22 (1993), 361-366). The regulatory elements may further comprise transcriptional and/or translational enhancers functional in plants cells. Furthermore, the regulatory elements may include transcription termination signals, such as a poly-A signal, which lead to the addition of a poly A tail to the transcript which may improve its stability.

In the case that a nucleic acid molecule according to the invention is expressed in sense orientation it is in principle possible to modify the coding sequence in such a way that the protein is located in any desired compartment of the plant cell. These include the nucleus, endoplasmatic reticulum, the vacuole, the mitochondria, the plastids, the apoplast, the cytoplasm etc. Since cyclin-dependent kinases the interacting component of the protein of the invention exert their effects in the cytoplasm and/or nucleus, corresponding signal sequences are preferred to direct the protein of the invention in the same compartment. Methods how to carry out this

modifications and signal sequences ensuring localization in a desired compartment are well known to the person skilled in the art.

Methods for the introduction of foreign DNA into plants are also well known in the art. These include, for example, the transformation of plant cells or tissues with T-DNA using *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*, the fusion of protoplasts, direct gene transfer (see, e.g., EP-A 164 575), injection, electroporation, biolistic methods like particle bombardment, pollen-mediated transformation, plant RNA virus-mediated transformation, liposome-mediated transformation, transformation using wounded or enzyme-degraded immature embryos, or wounded or enzyme-degraded embryogenic callus and other methods known in the art. The vectors used in the method of the invention may contain further functional elements, for example "left border"- and "right border"-sequences of the T-DNA of *Agrobacterium* which allow for stably integration into the plant genome. Furthermore, methods and vectors are known to the person skilled in the art which permit the generation of marker free transgenic plants, i.e. the selectable or scorable marker gene is lost at a certain stage of plant development or plant breeding. This can be achieved by, for example cotransformation (Lyznik, *Plant Mol. Biol.* 13 (1989), 151-161; Peng, *Plant Mol. Biol.* 27 (1995), 91-104) and/or by using systems which utilize enzymes capable of promoting homologous recombination in plants (see, e.g., WO97/08331; Bayley, *Plant Mol. Biol.* 18 (1992), 353-361; Lloyd, *Mol. Gen. Genet.* 242 (1994), 653-657; Maeser, *Mol. Gen. Genet.* 230 (1991), 170-176; Onouchi, *Nucl. Acids Res.* 19 (1991), 6373-6378). Methods for the preparation of appropriate vectors are described by, e.g., Sambrook (*Molecular Cloning; A Laboratory Manual*, 2nd Edition (1989), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Suitable strains of *Agrobacterium tumefaciens* and vectors as well as transformation of *Agrobacteria* and appropriate growth and selection media are well known to those skilled in the art and are described in the prior art (GV3101 (pMK90RK), Koncz, *Mol. Gen. Genet.* 204 (1986), 383-396; C58C1 (pGV 3850kan), Deblaere, *Nucl. Acid Res.* 13 (1985), 4777; Bevan, *Nucleic. Acid Res.* 12(1984), 8711; Koncz, *Proc. Natl. Acad. Sci. USA* 86 (1989), 8467-8471; Koncz, *Plant Mol. Biol.* 20 (1992),

963-976; Koncz, Specialized vectors for gene tagging and expression studies. In: Plant Molecular Biology Manual Vol 2, Gelvin and Schilperoort (Eds.), Dordrecht, The Netherlands: Kluwer Academic Publ. (1994), 1-22; EP-A-120 516; Hoekema: The Binary Plant Vector System, Offsetdrukkerij Kanter B.V., Alblasterdam (1985), Chapter V, Fraley, Crit. Rev. Plant. Sci., 4, 1-46; An, EMBO J. 4 (1985), 277-287). Although the use of *Agrobacterium tumefaciens* is preferred in the method of the invention, other *Agrobacterium* strains, such as *Agrobacterium rhizogenes*, may be used, for example if a phenotype conferred by said strain is desired.

Methods for the transformation using biolistic methods are well known to the person skilled in the art; see, e.g., Wan, Plant Physiol. 104 (1994), 37-48; Vasil, Bio/Technology 11 (1993), 1553-1558 and Christou (1996) Trends in Plant Science 1, 423-431. Microinjection can be performed as described in Potrykus and Spangenberg (eds.), Gene Transfer To Plants. Springer Verlag, Berlin, NY (1995).

The transformation of most dicotyledonous plants is possible with the methods described above. But also for the transformation of monocotyledonous plants several successful transformation techniques have been developed. These include the transformation using biolistic methods as, e.g., described above as well as protoplast transformation, electroporation of partially permeabilized cells, introduction of DNA using glass fibers, etc.

The term "transformation" as used herein, refers to the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for the transfer. The polynucleotide may be transiently or stably introduced into the host cell and may be maintained non-integrated, for example, as a plasmid, or alternatively, may be integrated into the host genome. The resulting transformed plant cell can then be used to regenerate a transformed plant in a manner known by a skilled person.

In general, the plants which can be modified according to the invention and which either show overexpression of a protein according to the invention or a reduction of the synthesis of such a protein can be derived from any desired plant species. They can be monocotyledonous plants or dicotyledonous plants, preferably they belong to plant species of interest in agriculture, wood culture or horticulture interest, such as crop plants (e.g. maize, rice, barley, wheat, rye, oats etc.), potatoes, oil producing

plants (e.g. oilseed rape, sunflower, pea nut, soy bean, etc.), cotton, sugar beet, sugar cane, leguminous plants (e.g. beans, peas etc.), wood producing plants, preferably trees, etc.

Thus, the present invention relates also to transgenic plant cells which contain stably integrated into the genome a nucleic acid molecule according to the invention linked to regulatory elements which allow for expression of the nucleic acid molecule in plant cells and wherein the nucleic acid molecule is foreign to the transgenic plant cell. For the meaning of foreign; see supra. Alternatively, a plant cell having (a) nucleic acid molecule(s) encoding a cyclin-dependent kinase inhibitor present in its genome can be used and modified such that said plant cell expresses the endogenous gene(s) corresponding to these nucleic acid molecules under the control of an heterologous promoter and/or enhancer elements. The introduction of the heterologous promoter and mentioned elements which do not naturally control the expression of a nucleic acid molecule encoding the above described protein using, e.g., gene targeting vectors can be done according to standard methods, see supra and, e.g., Hayashi, *Science* 258 (1992), 1350-1353; Fritze and Walden, *Gene activation by T-DNA tagging*. In *Methods in Molecular biology* 44 (Gartland, K.M.A. and Davey, M.R., eds). Totowa: Human Press (1995), 281-294) or transposon tagging (Chandlee, *Physiologia Plantarum* 78 (1990), 105-115). Suitable promoters and other regulatory elements such as enhancers include those mentioned hereinbefore.

The presence and expression of the nucleic acid molecule in the transgenic plant cells leads to the synthesis of a cyclin-dependent kinase inhibitor and leads to physiological and phenotypic changes in plants containing such cells.

Thus, the present invention also relates to transgenic plants and plant tissue comprising transgenic plant cells according to the invention. Due to the (over)expression of a cell cycle interacting protein of the invention, e.g., at developmental stages and/or in plant tissue in which they do not naturally occur these transgenic plants may show various physiological, developmental and/or morphological modifications in comparison to wild-type plants. For example, these

transgenic plants may display an altered cell elongation and/or for improved and/or disease resistance.

Therefore, part of this invention is the use of CKIs and the encoding DNA sequences to modulate plant cell division and/or growth in plant cells, plant tissues, plant organs and/or whole plants. To the scope of the invention also belongs a method to influence the activity of cyclin-dependent protein kinase in a plant cell by transforming the plant cell with a nucleic acid molecule according to the invention and/or manipulation of the expression of said molecule. More in particular using a nucleic acid molecule according to the invention, the disruption of plant cell division can be accomplished by interfering in the activity of cyclin-dependent protein kinases or their inhibitors. The latter goal may also be achieved, for example, with methods for reducing the amount of active cyclin-dependent kinase inhibitor.

Hence, the invention also relates to a transgenic plant cell which contains (stably integrated into the genome) a nucleic acid molecule according to the invention or part thereof, wherein the transcription and/or expression of the nucleic acid molecule or part thereof leads to reduction of the synthesis of a cyclin-dependent kinase inhibitor.

In a preferred embodiment, the reduction is achieved by an anti-sense, sense, ribozyme, co-suppression and/or dominant mutant effect.

"Antisense" and "antisense nucleotides" means DNA or RNA constructs which block the expression of the naturally occurring gene product.

The provision of the nucleic acid molecules according to the invention opens up the possibility to produce transgenic plant cells with a reduced level of the protein as described above and, thus, with a defect in the accumulation of a cyclin-dependent kinase inhibitor. Techniques how to achieve this are well known to the person skilled in the art. These include, for example, the expression of antisense-RNA, ribozymes, of molecules which combine antisense and ribozyme functions and/or of molecules which provide for a co-suppression effect; see also supra. When using the antisense approach for reduction of the amount of cyclin-dependent kinase inhibitor in plant cells, the nucleic acid molecule encoding the antisense-RNA is preferably of homologous origin with respect to the plant species used for transformation. However,

it is also possible to use nucleic acid molecules which display a high degree of homology to endogenously occurring nucleic acid molecules encoding a cyclin-dependent kinase inhibitor. In this case the homology is preferably higher than 80%, particularly higher than 90% and still more preferably higher than 95%.

The reduction of the synthesis of a protein according to the invention in the transgenic plant cells can result in an alteration in, e.g., cell division. In transgenic plants comprising such cells this can lead to various physiological, developmental and/or morphological changes.

Thus, the present invention also relates to transgenic plants comprising the above-described transgenic plant cells. These may show, for example, reduced or enhanced growth characteristics.

The present invention also relates to cultured plant tissues comprising transgenic plant cells as described above which either show overexpression of a protein according to the invention or a reduction in synthesis of such a protein.

Any transformed plant obtained according to the invention can be used in a conventional breeding scheme or in *in vitro* plant propagation to produce more transformed plants with the same characteristics and/or can be used to introduce the same characteristic in other varieties of the same or related species. Such plants are also part of the invention. Seeds obtained from the transformed plants genetically also contain the same characteristic and are part of the invention. As mentioned before, the present invention is in principle applicable to any plant and crop that can be transformed with any of the transformation method known to those skilled in the art and includes for instance corn, wheat, barley, rice, oilseed crops, cotton, tree species, sugar beet, cassava, tomato, potato, numerous other vegetables, fruits.

In yet another aspect, the invention also relates to harvestable parts and to propagation material of the transgenic plants according to the invention which either contain transgenic plant cells expressing a nucleic acid molecule according to the

invention or which contain cells which show a reduced level of the described protein. Harvestable parts can be in principle any useful parts of a plant, for example, flowers, pollen, seedlings, tubers, leaves, stems, fruit, seeds, roots etc. Propagation material includes, for example, seeds, fruits, cuttings, seedlings, tubers, rootstocks etc.

As mentioned above, the cyclin-dependent kinase inhibitors of the invention display distinct expression patterns in plants and cell suspension. Thus, the regulatory sequences that naturally drive the expression of the above described cyclin-dependent kinase inhibitors may prove useful for the expression of heterologous DNA sequences in certain plant tissues and/or at different developmental stages in plant development.

Accordingly, in a further aspect the present invention relates to a regulatory sequence of a promoter naturally regulating the expression of a nucleic acid molecule of the invention described above or of a nucleic acid molecule homologous to a nucleic acid molecule of the invention. The expression pattern of CKI genes has been studied in detail in accordance with the present invention and is summarized in Example 8, 9 and in particular in Example 13. With methods well known in the art it is possible to isolate the regulatory sequences of the promoters that naturally regulate the expression of the above-described DNA sequences. For example, using the CKI genes as probes a genomic library consisting of plant genomic DNA cloned into phage or bacterial vectors can be screened by a person skilled in the art. Such a library consists e.g. of genomic DNA prepared from seedlings, fractionized in fragments ranging from 5 kb to 50 kb, cloned into the lambda GEM11 (Promega) phages. Phages hybridizing with the probes can be purified. From the purified phages DNA can be extracted and sequenced. Having isolated the genomic sequences corresponding to the genes encoding the above-described cyclin-dependent kinase inhibitors, it is possible to fuse heterologous DNA sequences to these promoters or their regulatory sequences via transcriptional or translational fusions well known to the person skilled in the art. In order to identify the regulatory sequences and specific elements of the CKI genes, 5'-upstream genomic fragments can be cloned in front of marker genes such as luc, gfp or the

GUS coding region and the resulting chimeric genes can be introduced by means of *Agrobacterium tumefaciens* mediated gene transfer into plants or transfected into plant cells or plant tissue for transient expression. The expression pattern observed in the transgenic plants or transfected plant cells containing the marker gene under the control of the regulatory sequences of the invention reveal the boundaries of the promoter and its regulatory sequences. Preferably, said regulatory sequence is capable of conferring expression of a heterologous DNA sequence in

- (a) young root meristems, pericycle cells in the vascular tissue, shoot apical meristem, surface and tip of young leaves, epidermis of the stem in young seedlings, tapetal layer of the anthers in pollen grains, flower buds and mature ovaries, embryos at the globular, heart and torpedo stages, embryonic root;
- (b) root and shoot apical meristems, young differentiating leaves, flower buds and young flowers, ovary wall, funiculus, ovules and pollen grains, embryo at the globular stage, embryonic root; or
- (c) main and lateral root meristems and shoot apical meristems, vascular tissue, pericycle, mature ovaries, globular and heart embryonic root.

In context with the present invention, the term "regulatory sequence" refers to sequences which influence the specificity and/or level of expression, for example in the sense that they confer cell and/or tissue specificity; see *supra*. Such regions can be located upstream of the transcription initiation site, but can also be located downstream of it, e.g., in transcribed but not translated leader sequences.

The term "promoter", within the meaning of the present invention refers to nucleotide sequences necessary for transcription initiation, i.e. RNA polymerase binding, and may also include, for example, the TATA box.

The term "nucleic acid molecule homologous to a nucleic acid molecule of the invention", as used herein includes promoter regions and regulatory sequences of other CKI genes, such as the gene encoding the CKI1 protein as well as genes from other species, for example, maize, alfalfa, potato, sorghum, millet, coix, barley, wheat and rice which are homologous to the CKI genes and which display substantially the same expression pattern. Such promoters are characterized by

their capability of conferring expression of a heterologous DNA sequence in root meristems and other tissues mentioned above.

Thus, according to the present invention, regulatory sequences from any species can be used that are functionally homologous to the regulatory sequences of the promoter of the above defined CKI specific nucleic acid molecules, or promoters of genes that display an identical or similar pattern of expression, in the sense of being expressed in the above-mentioned tissues and cells. However, the expression conferred by the regulatory sequences of the invention may not be limited to, for example, root meristem cells but can include or be restricted to, for example, subdomains of meristems. The particular expression pattern may also depend on the plant/vector system employed. However, expression of heterologous DNA sequences driven by the regulatory sequences of the invention predominantly occurs in the root meristem unless certain elements of the regulatory sequences of the invention, were taken and designed by the person skilled in the art to control the expression of a heterologous DNA sequence in other cell types.

It is also immediately evident to the person skilled in the art that further regulatory elements may be added to the regulatory sequences of the invention. For example, transcriptional enhancers and/or sequences which allow for induced expression of the regulatory sequences of the invention may be employed. A suitable inducible system is for example tetracycline-regulated gene expression as described, e.g., by Gatz, *supra*.

The regulatory sequence of the invention may be derived from the CKI genes of *Arabidopsis thaliana* or alfalfa although other plants may be suitable sources for such regulatory sequences as well.

Usually, said regulatory sequence is part of a recombinant DNA molecule. In a preferred embodiment of the present invention, the regulatory sequence in the recombinant DNA molecule is operatively linked to a heterologous DNA sequence. The term heterologous with respect to the DNA sequence being operatively linked to the regulatory sequence of the invention means that said DNA sequence is not

naturally linked to the regulatory sequence of the invention. Expression of said heterologous DNA sequence comprises transcription of the DNA sequence, preferably into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably plant cells, are well known to those skilled in the art. They usually comprise poly-A signals ensuring termination of transcription and stabilization of the transcript, see also supra. Additional regulatory elements may include transcriptional as well as translational enhancers; see supra.

In a preferred embodiment, the heterologous DNA sequence of the above-described recombinant DNA molecules encodes a peptide, protein, antisense RNA, sense RNA and/or ribozyme. The recombinant DNA molecule of the invention can be used alone or as part of a vector to express heterologous DNA sequences, which, e.g., encode proteins for, e.g., the control of disease resistance, modulation of nutrition value or diagnostics of CKI related gene expression. The recombinant DNA molecule or vector containing the DNA sequence encoding a protein of interest is introduced into the cells which in turn produce the protein of interest. For example, the regulatory sequences of the invention can be operatively linked to sequences encoding Barstar and Barnase, respectively, for use in the production of male and female sterility in plants.

On the other hand, said protein can be a scorable marker, e.g., luciferase, green fluorescent protein or β -galactosidase. This embodiment is particularly useful for simple and rapid screening methods for compounds and substances described herein below capable of modulating CKI specific gene expression. For example, a cell suspension can be cultured in the presence and absence of a candidate compound in order to determine whether the compound affects the expression of genes which are under the control of regulatory sequences of the invention, which can be measured, e.g., by monitoring the expression of the above-mentioned marker. It is also immediately evident to those skilled in the art that other marker genes may be employed as well, encoding, for example, a selectable marker which provides for the direct selection of compounds which induce or inhibit the expression of said marker.

The regulatory sequences of the invention may also be used in methods of antisense approaches. The antisense RNA may be a short (generally at least 10, preferably at least 14 nucleotides, and optionally up to 100 or more nucleotides) nucleotide sequence formulated to be complementary to a portion of a specific mRNA sequence and/or DNA sequence of the gene of interest. Standard methods relating to antisense technology have been described; see, e.g., Klann, *Plant Physiol.* 112 (1996), 1321-1330. Following transcription of the DNA sequence into antisense RNA, the antisense RNA binds to its target sequence within a cell, thereby inhibiting translation of the mRNA and down-regulating expression of the protein encoded by the mRNA. Thus, in a further embodiment, the invention relates to nucleic acid molecules of at least 15 nucleotides in length hybridizing specifically with a regulatory sequence as described above or with a complementary strand thereof. For the possible applications of such nucleic acid molecules, see *supra*.

The present invention also relates to vectors, particularly plasmids, cosmids, viruses and bacteriophages used conventionally in genetic engineering that comprise a recombinant DNA molecule of the invention. Preferably, said vector is an expression vector and/or a vector further comprising a selection marker for plants. For example of suitable selector markers, see *supra*. Methods which are well known to those skilled in the art can be used to construct recombinant vectors; see, for example, the techniques described in Sambrook, *Molecular Cloning A Laboratory Manual*, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the recombinant DNA molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

The present invention furthermore relates to host cells transformed with a regulatory sequence, a DNA molecule or vector of the invention. Said host cell may be a prokaryotic or eukaryotic cell; see *supra*.

In a further preferred embodiment, the present invention provides for a method for the production of transgenic plants, plant cells or plant tissue comprising the

introduction of a nucleic acid molecule, recombinant DNA molecule or vector of the invention into the genome of said plant, plant cell or plant tissue. For the expression of the heterologous DNA sequence under the control of the regulatory sequence according to the invention in plant cells, further regulatory sequences such as poly A tail may be fused, preferably 3' to the heterologous DNA sequence, see also supra. Further possibilities might be to add Matrix Attachment Sites at the borders of the transgene to act as "delimiters" and insulate against methylation spread from nearby heterochromatic sequences. Methods for the introduction of foreign DNA into plants, plant cells and plant tissue are described above.

Thus, the present invention relates also to transgenic plant cells which contain stably integrated into the genome a recombinant DNA molecule or vector according to the invention.

Furthermore, the present invention also relates to transgenic plants and plant tissue comprising the above-described transgenic plant cells. These plants may show, for example, increased disease resistance.

In yet another aspect the invention also relates to harvestable parts and to propagation material of the transgenic plants according to the invention which contain transgenic plant cells described above. Harvestable parts and propagation material can be in principle any useful part of a plant; see supra.

With the regulatory sequences of the invention, it will be possible to study *in vivo* CKI specific gene expression. Furthermore, since CKI specific gene expression has different patterns in different stages of physiological and pathological conditions, it is now possible to determine further regulatory sequences which may be important for the up- or down-regulation of CKI gene expression, for example in response to ions or elicitors. In addition, it is now possible to *in vivo* study mutations which affect different functional or regulatory aspects of specific gene expression in the cell cycle.

The *in vivo* studies referred to above will be suitable to further broaden the knowledge on the mechanisms involved in the control of the cell cycle. To date nothing is known about the activity, nature or mode of action of CKIs in the cell cycle or about their role during plant development. Expression of heterologous genes or antisense RNA under the control of the regulatory sequence of the present invention in plants and plant cells may allow the understanding of the function of each of these proteins in the plant.

The present invention further relates to a method for the identification of an activator or inhibitor of genes encoding cyclin-dependent kinase inhibitors comprising the steps of:

- (a) providing a plant, plant cell, or plant tissue comprising a recombinant DNA molecule comprising a readout system operatively linked to a regulatory sequence of the invention;
- (b) culturing said plant cell or tissue or maintaining said plant in the presence of a compound or a sample comprising a plurality of compounds under conditions which permit expression of said readout system;
- (c) identifying or verifying a sample and compound, respectively, which leads to suppression or activation and/or enhancement of expression of said readout system in said plant, plant cell, or plant tissue.

The present invention further relates to a method for identifying and obtaining an activator or inhibitor of cyclin-dependent kinase inhibitors comprising the steps of:

- (a) combining a compound to be screened with a reaction mixture containing the protein of the invention and a readout system capable of interacting with the protein under suitable conditions;
- (b) maintaining said reaction mixture in the presence of the compound or a sample comprising a plurality of compounds under conditions which permit interaction of the protein with said readout system;
- (c) identifying or verifying a sample and compound, respectively, which leads to suppression or activation of the readout system.

The term "read out system" in context with the present invention means a DNA sequence which upon transcription and/or expression in a cell, tissue or organism provides for a scorable and/or selectable phenotype. Such read out systems are well known to those skilled in the art and comprise, for example, recombinant DNA molecules and marker genes as described above and in the appended example.

The term "plurality of compounds" in a method of the invention is to be understood as a plurality of substances which may or may not be identical.

Said compound or plurality of compounds may be comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compound(s) may be known in the art but hitherto not known to be capable of suppressing or activating cell cycle interacting proteins. The reaction mixture may be a cell free extract or may comprise a cell or tissue culture. Suitable set ups for the method of the invention are known to the person skilled in the art and are, for example, generally described in Alberts et al., *Molecular Biology of the Cell*, third edition (1994), in particular Chapter 17. The plurality of compounds may be, e.g., added to the reaction mixture, culture medium or injected into the cell.

If a sample containing a compound or a plurality of compounds is identified in the method of the invention, then it is either possible to isolate the compound from the original sample identified as containing the compound capable of suppressing or activating cyclin-dependent kinase inhibitors, or one can further subdivide the original sample, for example, if it consists of a plurality of different compounds, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. Depending on the complexity of the samples, the steps described above can be performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances of similar chemical and/or physical properties, and most preferably said substances are identical. Preferably, the compound identified according to the above described method or its derivative is further formulated in a form suitable for the application in plant breeding or plant cell and tissue culture.

The compounds which can be tested and identified according to a method of the invention may be expression libraries, e.g., cDNA expression libraries, peptides, proteins, nucleic acids, antibodies, small organic compounds, hormones, peptidomimetics, PNAs or the like (Milner, *Nature Medicine* 1 (1995), 879-880; Hupp, *Cell* 83 (1995), 237-245; Gibbs, *Cell* 79 (1994), 193-198 and references cited supra). Furthermore, genes encoding a putative regulator of a cyclin-dependent kinase inhibitor and/or which exert their effects up- or downstream the cell cycle interacting protein of the invention may be identified using, for example, insertion mutagenesis using, for example, gene targeting vectors known in the art (see, e.g., Hayashi, *Science* 258 (1992), 1350-1353; Fritze and Walden, *Gene activation by T-DNA tagging*. In *Methods in Molecular biology* 44 (Gartland, K.M.A. and Davey, M.R., eds). Totowa: Human Press (1995), 281-294) or transposon tagging (Chandlee, *Physiologia Plantarum* 78 (1990), 105-115). Said compounds can also be functional derivatives or analogues of known inhibitors or activators. Methods for the preparation of chemical derivatives and analogues are well known to those skilled in the art and are described in, for example, Beilstein, *Handbook of Organic Chemistry*, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and *Organic Synthesis*, Wiley, New York, USA. Furthermore, said derivatives and analogues can be tested for their effects according to methods known in the art. Furthermore, peptidomimetics and/or computer aided design of appropriate derivatives and analogues can be used, for example, according to the methods described above. The cell or tissue that may be employed in the method of the invention preferably is a host cell, plant cell or plant tissue of the invention described in the embodiments hereinbefore.

Determining whether a compound is capable of suppressing or activating cell cycle interacting proteins can be done, for example, by monitoring DNA duplication and cell division. It can further be done by monitoring the phenotypic characteristics of the cell of the invention contacted with the compounds and compare it to that of wild-type plants. In an additional embodiment, said characteristics may be compared to that of a cell contacted with a compound which is either known to be capable or incapable of suppressing or activating cell cycle interacting proteins.

The inhibitor or activator identified by the above-described method may prove useful as a herbicide, pesticide and/or as a plant growth regulator. Thus, in a further embodiment the invention relates to a compound obtained or identified according to the method of the invention said compound being an activator of a cyclin-dependent kinase inhibitor or an inhibitor of a cyclin-dependent kinase inhibitor.

Such useful compounds can be for example transacting factors which bind to the cyclin-dependent kinase inhibitor of the invention. Identification of transacting factors can be carried out using standard methods in the art (see, e.g., Sambrook, *supra*, and Ausubel, *supra*). To determine whether a protein binds to the protein of the invention, standard native gel-shift analyses can be carried out. In order to identify a transacting factor which binds to the protein of the invention, the protein of the invention can be used as an affinity reagent in standard protein purification methods, or as a probe for screening an expression library. Once the transacting factor is identified, modulation of its binding to the cyclin-dependent kinase inhibitor of the invention can be pursued, beginning with, for example, screening for inhibitors against the binding of the transacting factor to the protein of the present invention. Activation or repression of cyclin-dependent kinase inhibitor could then be achieved in plants by applying of the transacting factor (or its inhibitor) or the gene encoding it, e.g. in a vector for transgenic plants. In addition, if the active form of the transacting factor is a dimer, dominant-negative mutants of the transacting factor could be made in order to inhibit its activity. Furthermore, upon identification of the transacting factor, further components in the pathway leading to activation (e.g. signal transduction) or repression of a gene involved in the control of cell cycle then can be identified. Modulation of the activities of these components can then be pursued, in order to develop additional drugs and methods for modulating the cell cycle in animals and plants.

The invention also relates to a diagnostic composition comprising at least one of the aforementioned nucleic acid molecules, vectors, proteins, antibodies, regulatory

sequences, recombinant DNA molecules, or compounds and optionally suitable means for detection.

Said diagnostic compositions may be used for methods for detecting expression of cyclin-dependent kinase inhibitors by detecting the presence of the corresponding mRNA which comprises isolation of mRNA from a cell and contacting the mRNA so obtained with a probe comprising a nucleic acid probe as described above under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of the protein in the cell. Further methods of detecting the presence of a protein according to the present invention comprises immunotechniques well known in the art, for example enzyme linked immunosorbent assay. Furthermore, it is possible to use the nucleic acid molecules according to the invention as molecular markers in plant breeding.

The person skilled in the art can use proteins according to the invention from other organisms such as yeast and animals to influence cell division progression in those other organisms such as mammals or insects. In a preferred embodiment one or more DNA sequences, vectors or proteins of the invention or the above-described antibody or compound are, for instance, used to specifically interfere in the disruption of the expression levels of genes involved in G1/S and/or G2/M transition in the cell cycle process in transformed plants, particularly :

- in the complete plant
- in selected plant organs, tissues or cell types
- under specific environmental conditions, including abiotic stress such as cold, heat, drought or salt stress or biotic stress such as pathogen attack
- during specific developmental stages.

Specifically the plant cell division rate and/or the inhibition of a plant cell division can be influenced by (partial) elimination of a gene or reducing the expression of a gene encoding a protein according to the invention. Said plant cell division rate and/or the inhibition of a plant cell division can also be influenced by eliminating or inhibiting the activity of the protein according to the invention by using for instance antibodies directed against said protein. As a result of said elimination or reduction greater organisms or specific organs or tissues can be obtained; greater in volume and in

mass too. Furthermore inhibition of cell division by various adverse environmental conditions such as drought, high salt content, chilling and the like can be delayed or prevented by reduction of said expression of a gene according to the invention. The division rate of a plant cell can also be influenced in a transformed plant by overexpression of a sequence according to the invention. Said transformed plant can be obtained by transforming a plant cell with a gene encoding a polypeptide concerned or fragment thereof alone or in combination, whereas the plant cell may belong to a monocotyledonous or dicotyledonous plant. For this purpose tissue specific promoters, in one construct or being present as a separate construct in addition to the sequence concerned, can be used. Therefore an important aspect of the current invention is a method to modify plant architecture by overproduction or reduction of expression of a sequence according to the invention under the control of a tissue, cell or organ specific promoter. Another aspect of the present invention is a method to modify the growth inhibition of plants caused by environmental stress conditions above mentioned by appropriate use of sequences according to the invention. Surprisingly using a polypeptide or fragment thereof according to the invention or using antisense RNA or any method to reduce the expression of the gene according to the invention, cell division in the meristem of both main and lateral roots, shoot apical or the vascular tissue of a plant can be manipulated. Furthermore any of the DNA sequences of the invention as well as those encoding CDK1 can be used to manipulate (reduce or enhance) the level of endopolyploidy and thereby increasing the storage capacity of, for example, endosperm cells.

Another aspect of the current invention is that one or more DNA sequences, vectors or proteins, regulatory sequences or recombinant DNA molecules of the invention or the above-described antibody or compound can be used to modulate, for instance, endoreduplication in storage cells, storage tissues and/or storage organs of plants or parts thereof. The term "endoreduplication" means recurrent DNA replication without consequent mitosis and cytokinesis.

Preferred target storage organs and parts thereof for the modulation of endoreduplication are, for instance, seeds (such as from cereals, oilseed crops), roots (such as in sugar beet), tubers (such as in potato) and fruits (such as in

vegetables and fruit species). Furthermore it is expected that increased endoreduplication in storage organs and parts thereof correlates with enhanced storage capacity and as such with improved yield. In yet another embodiment of the invention, a plant with modulated endoreduplication in the whole plant or parts thereof can be obtained from a single plant cell by transforming the cell, in a manner known to the skilled person, with the above-described means.

In view of the foregoing, the present invention also relates to the use of a DNA sequence, vector, protein, antibody, regulatory sequences, recombinant DNA molecule, nucleic acid molecules or compound of the invention for modulating plant cell cycle, plant cell division and/or growth, for influencing the activity of cyclin-dependent protein kinase, for disrupting plant cell division by influencing the presence or absence or by interfering in the expression of a cyclin-dependent protein kinase inhibitor, for modifying growth inhibition of plants caused by environmental stress conditions, for inducing male or female sterility, for influencing cell division progression in a host as defined above or for use in a screening method for the identification of inhibitors or activators of cell cycle proteins. Beside the above described possibilities to use the nucleic acid molecules according to the invention for the genetic engineering of plants with modified characteristics and their use to identify homologous molecules, the described nucleic acid molecules may also be used for several other applications, for example, for the identification of nucleic acid molecules which encode proteins which interact with the cell cycle proteins described above. This can be achieved by assays well known in the art such as those described above and also included, for example, as described in Scofield (Science 274 (1996), 2063-2065) by use of the so-called yeast "two-hybrid system"; see also the appended examples. In this system the protein encoded by the nucleic acid molecules according to the invention or a smaller part thereof is linked to the DNA-binding domain of the GAL4 transcription factor. A yeast strain expressing this fusion protein and comprising a lacZ reporter gene driven by an appropriate promoter, which is recognized by the GAL4 transcription factor, is transformed with a library of cDNAs which will express plant proteins or peptides thereof fused to an activation domain. Thus, if a peptide encoded by one of the cDNAs is able to interact with the fusion peptide comprising a peptide of a protein of the invention, the complex is able to

direct expression of the reporter gene. In this way the nucleic acid molecules according to the invention and the encoded peptide can be used to identify peptides and proteins interacting with cell cycle interacting proteins. It is apparent to the person skilled in the art that this and similar systems may then further be exploited for the identification of inhibitors of the binding of the interacting proteins.

Other methods for identifying compounds which interact with the proteins according to the invention or nucleic acid molecules encoding such molecules are, for example, the in vitro screening with the phage display system as well as filter binding assays or "real time" measuring of interaction using, for example, the BIAcore apparatus (Pharmacia); see references cited supra.

Furthermore, it is possible to use the nucleic acid molecules according to the invention as molecular markers in plant breeding. Moreover, the overexpression of nucleic acid molecules according to the invention may be useful for the alteration or modification of plant/pathogene interaction. The term "pathogene" includes, for example, bacteria, viruses and fungi as well as protozoa.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. Further literature concerning any one of the methods, material, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under <http://www.ncbi.nlm.nih.gov/PubMed/medline.html>. Further databases and addresses, such as <http://www.ncbi.nlm.nih.gov/>, <http://www.infobiogen.fr/>, http://www.fmi.ch/biology/research_tools.html, <http://www.tigr.org/>, are known to the person skilled in the art and can also be obtained using, e.g., <http://www.lycos.com>. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.

The present invention is further described by reference to the following non-limiting figures and examples.

The Figures show:

Figure 1: Sequence alignment of the *Arabidopsis thaliana* cyclin-dependent kinase inhibitors FL39, FL66, FL67, ICKI (accession number AC003040); the *Medicago sativa* cyclin-dependent kinase inhibitor ALFCDKI, and the *Chenopodium rubrum* cyclin-dependent kinase CrCKI (accession number AJ002173). Alignment was obtained using the PILEUP program (from the GCG 9.1 package) using the parameters Gap weight = 4 and Length weight = 0.

The Examples illustrate the invention:

Unless stated otherwise in the examples, all recombinant DNA techniques are performed according to protocols as described in Sambrook et al. (1989), Molecular Cloning : A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY or in Volumes 1 and 2 of Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfax (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

Example 1: Identification of putative cyclin-dependent kinase inhibitors

For the identification of CKIs a two hybrid system based on GAL4 recognition sites to regulate the expression of both *his3* and *lacZ* reporter genes was used to identify CDC2aAt-interacting of proteins. The bait used for the two-hybrid screening was constructed by inserting the *CDC2aAt* coding region into the pGBT9 vector (Clontech). The insert was created by PCR using the *CDC2aAt* cDNA as template.

Primers were designed to incorporate *EcoRI* restriction enzyme sites. The primers used were 5'-CGAGATCTGAATTCATGGATCAGTA-3' (SEQ ID NO: 7) and 5'-CGAGATCTGAATTCCTAAGGCATGCC-3' (SEQ ID NO: 8). The PCR fragment was cut with *EcoRI* and cloned into the *EcoRI* site of pGBT9, resulting in the pGBTCDC2A plasmid. For the screening a GAL4 activation domain cDNA fusion library was used constructed from *Arabidopsis thaliana* cell suspension cultures. This library was constructed using RNA isolated from cells harvested at 20 hours, 3, 7 and 10 days after dilution of the culture in new medium. These time point correspondent to cells from the early exponential growth phase to the late stationary phase. mRNA was prepared using Dynabeads oligo(dT)₂₅ according to the manufacturer's instructions (Dyna). The GAL4 activation domain cDNA fusion library was generated using the HybriZAP™ vector purchased with the HybriZAP™ Two-Hybrid cDNA Gigapack cloning Kit (Stratagene) following the manufacturer's instructions. The resulting library contained approximately $3 \cdot 10^6$ independent plaque-forming units, with an average insert size of 1 Kb.

For the screening a 1-liter culture of the *Saccharomyces cerevisiae* strain HF7c (*MAT_a*, *ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538 LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3 URA3::GAL4_{17mers(3x)}-CyC1_{TATA}-LacZ*) was cotransformed with 400 µg pGBTCDC2A, 500 µg DNA of the library, and 40 mg salmon sperm carrier DNA using the lithium acetate method (Gietz *et al.* 1992, Nucleic Acids Res. 20, 1425). To estimate the number of independent cotransformants, 1/1000 of the transformation mix was plated on Leu⁻ and Trp⁻ medium. The rest of the transformation mix was plated on medium to select for histidine prototrophy (Trp⁺, Leu⁺, His⁺). Of a total of approximately 1.2×10^7 independent transformants 1200 colonies grew after 3 days of incubation at 30°C. The colonies larger than 2 mm were streaked on histidine-lacking medium supplemented with 10 mM 3-amino-1,2,4-triazole (Sigma). Two-hundred-fifty colonies capable of growing under these conditions were tested for β-galactosidase activity as described (Breedon and Nasmyth 1995, Cold Spring Harbor Symp. Quant. Biol. 50, p643-650), and 153 turned out to be His⁺ and LacZ⁺. Plasmid DNA was prepared from the positive clones and sequenced.

The plasmids pGADLDV39, pGADLDV66, and pGADLDV159 contained a protein (designated LDV39, LDV66, and LDV159, respectively) of which the last 23 amino-acids showed significant homology to the human CKIs p21^{cip1} and p27^{kip1}. The LDV159 clone was identical to ICK1 (GenBank accession number U94772 as published by Wang in Nature 386 (1997), 451-452). The two other clones were novel and encoded proteins only distantly related to ICK1 (Table 1). The LDV39 gene was 622 bp long, consisting of 423 bp coding region and 199 bp 3' UTR (excluding the poly-A tail). The LDV66 gene was 611 bp long, consisting of 379 bp coding region and 232 bp 3' UTR (excluding the poly-A tail).

The specificity of the interaction between LDV39, LDV66, and LDV159 was verified by the retransformation of yeast with pGBTCDC2A and pGADLDV39/pGADLDV66/pGADLDV159. As controls, pGBTCDC2A was cotransformed with a vector containing only the GAL4 activation domain (pGAD424); and the pGADLDV39/pGADLDV66/pGADLDV159 vectors were cotransformed with a plasmid containing only the GAL4 DNA binding domain (pGBT9). Transformants were plated on medium with or without histidine. Only transformants containing both pGBTCDC2A and pGADLDV39, pGADLDV66, or pGADLDV159 were able to grow in the absence of histidine.

Example 2: LDV66, LDV39 and LDV159 bind CDC2aAt, not CDC2bAt

The pGBTCDC2B vector encoding a fusion protein between the C-terminus of the GAL4 DNA-binding domain and CDC2bAt was constructed by cloning the full length coding region of *CDC2bAt* into the pGBT9 vector. pGBTCDC2B was transformed with pGADLDV66/pGADLDV39/pGADLDV159 in the HF7c yeast and cotransformants were plated on medium with or without histidine. As control, pGBTCDC2A was transformed with pGADLDV66/pGADLDV39/pGADLDV159. In contrast to the transformants containing the pGBTCDC2A vector were cotransformants containing the pGBTCDC2B vector unable to grow in the absence of histidine. This demonstrates that the LDV66, LDV39, LDV159 proteins associate with CDC2aAt but not with CDC2bAt.

Example 3: Isolation of FL39 and FL66 sequences

Since the *LDV39* and *LDV66* clones encode partial proteins, lacking their amino-terminal part, a flower cDNA library obtained from the ABRC stock centre (library stock number CD4-6) was screened. In total 50.000 plaque forming units were hybridised using a fluorescein-labelled *LDV39* or *LDV66* probe according to the manufacturer's protocol (Amersham) using a hybridisation temperature of 60°C. After 16 hours hybridisation the filters were washed for 15 min using 2X SSC; 0.1X SDS, and 15 min using 1X SSC; 0.1X SDS. The signals were detected using the CDP-star detection module according to the manufacturer's protocol (Amersham). The signals were revealed by autoradiography. For both genes only one positive signal was identified among the 50.000 phages, suggesting low mRNA levels of *LDV39* and *LDV66* in flowers. Phages corresponding to the positive signals were eluted from gel and purified by two additional hybridisation rounds, using 1.000 and 50 plaque forming units in the second and third hybridisation round, respectively. The hybridisation conditions were similar as those described above. After pure phages were obtained, DNA was extracted and sequenced. The positive clones were denominated *FL39* and *FL66*, corresponding to longer clones of *LDV39* and *LDV66*, respectively.

The *FL39* clone is 932 bp long and contains an ORF encoding a protein of 209 amino acids with a calculated molecular mass of 24 kDa. In its 3' UTR a polyadenylation signal can be recognised. The amino-terminal part of the *FL39* protein contains a repeated motif of 11 amino acids (VRRRD/ExxxVEE; SEQ ID NO: 33). This motif is not found in any other protein in the databanks and its significance is unknown. The *FL39* protein also contains a putative nuclear localization signal (amino acids 23-26) and a PEST-rich region (amino acids 71-98; PESTFIND score +15.5). These sequences, rich in proline, glutamic acid, serine and proline, are characteristically present in unstable proteins (Rogers et al., 1986, Science 234, 364-368).

The *FL66* sequence does not contain an in frame stopcodon, and may therefore not be full length. The *FL66* clone is 875 bp long and bears an ORF of 216 amino acids,

encoding a protein of 24 kD. No nuclear localization signal or PEST domains are present.

The genomic organisation of the *FL39*, *FL66* and *LDV159* clones was tested by DNA gel blot analysis. *A. thaliana* C24 DNA digested with three different restriction enzymes was probed with fluorescein-labelled prepared from the *LDV159*, *FL39*, or *FL66* sequences according to the manufacturer's protocol (Amersham). Hybridisations were performed at 60°C. After 16 hours hybridisation the membranes were washed for 15 min using 2X SSC; 0.1X SDS, and 15 min using 1X SSC; 0.1X SDS. The signals were detected using the CDP-star detection module according to the manufacturer's protocol (Amersham). The signals were revealed by autoradiography. For *LDV159* and *FL39*, only one hybridisation band was noticed for every digest. For *FL66* an additional weak band was observed. The low intensity bands did not correspond with any of the bands found for *LDV159* or *FL39*, suggesting the presence of an additional *FL66* related gene. We conclude that there are at least four different CKI proteins present in *A. thaliana*.

Example 4: The *Arabidopsis thaliana* CKIs bind exclusively to CDC2aAt *in vivo*

The binding specificity of the *FL39* and *FL66* proteins towards CDC2aAt and CDC2bAt was studied using the two-hybrid system. The *FL39* and *FL66* coding regions were cloned in frame with the GAL4 activation-domain in the pGAD424 vector (Clontech). The *FL39* coding region was amplified using the 5'-GGGAATCCATGGGCGGCGGTTAGGAGAAG-3' (SEQ ID NO: 9) and 5'-GGCGGATCCCGTCTTCTTCATGGATTC-3' (SEQ ID NO: 10) primers. The *FL66* coding region was amplified using the 5'-GGCGAATCCATGGAAGTCTCTAAAGCAAC-3' (SEQ ID NO: 11) and 5'-GGCGGATCCTTTTGAAGTTCATGGTTTGAC-3' (SEQ ID NO: 12) primers. The *FL66* amplified coding sequence encloses a protein starting at the methionine at amino-acid position 11, therefore not including the first 10 amino-acids encoded by the *FL66* clone. The PCR fragments were cut with *EcoR1* and *BamH1* and cloned into the *EcoR1* and *BamH1* sites of pGAD424, resulting in the pGADFL39 and

pGADFL66 clones. These plasmids were transformed into the HF7c yeast in combination with pGBTCDC2A or pGBTCDC2B. The pGBTCDC2B plasmid, encoding a fusion protein between the C-terminus of the GAL4 DNA-binding domain and CDC2bAt was obtained by cloning the full length coding of CDC2bAt into the pGBT9 vector (Clontech).

In contrast to the transformants containing the pGBTCDC2A vector were the transformants containing the pGBTCDC2B vector unable to grow in the absence of histidine. This demonstrates that the FL39 and FL66 proteins exclusively associate with CDC2aAt.

Example 5: Generation of FL39 and FL66 specific antibodies

To obtain sufficient amount of FL66 and FL39 proteins for immunization, the *FL39* and *FL66* coding sequences were cloned into pET vectors. The genes cloned in these vectors are expressed under the control of the strong inducible T7 promoter in *Escherichia coli* (Studier et al., 1986, J. Mol.Biol., 189, p113-130). The coding region of *FL39* and *FL66* were amplified by PCR technique. The FL66 amplified coding sequence encloses a protein starting at the methionine at amino-acid position 11, therefore not including the first 10 amino-acids encoded by the *FL66* clone. Primers used to amplify *FL39* were 5'-TAGGAGCATATGGCGGCGG-3' (SEQ ID NO: 29) and 5'-ATCATCGAATTCTTCATGGATTC-3' (SEQ ID NO: 30). Primers used to amplify *FL66* were 5'-ATATCAGCGCCATGGAAGTC-3' (SEQ ID NO: 31) and 5'-GGAGCTGGATCCTTTTGAATTTCATGG-3' (SEQ ID NO: 32).

The obtained *FL39* PCR fragment was purified, and cut with *NdeI* and *EcoRI* restriction enzymes. This fragment was cloned into the *NdeI* and *EcoRI* sites of pET derivative pRK172 (McLeod et al., 1987, EMBO J. 6, p729-736). The obtained *FL66* PCR fragment was purified, cut with *NcoI* and *BamHI* and cloned into the *NcoI* and *BamHI* sites of pET21d. FL66pET21d was transformed in *E. coli* BL21(DE3). FL39pRK172 was co-transformed in *E.coli* BL21(DE3) with pSBETa (Schenk et al., 1995, Biotechniques 19, p 196-200). PSBETa encoded the tRNA^{UCU} that is a low abundant tRNA in *E. coli*, corresponding to codons AGG and AGA (arginine).

Because of the presence of a AGG AGA AGA sequence (Arg 5, Arg 6, Arg 7) at the beginning of *FL39* coding sequence, an increase of the tRNA^{UCU} pool of *E. coli* is necessary for the translation of *FL39*.

The *FL66pET21d/BL21(DE3)* and *FL39pRK172, pSBETa/ BL21(DE3)* *E. coli* recombinant strains were grown in LB medium, supplemented respectively with 50 µg/ml ampicilline and 50 µg/ml ampicilline; 25 µg/ml kanamycine. The cells were grown at 37°C until the density of the culture reached an $A_{600nm} = 0.7$. At this time point, 0.4 mM IPTG was added to induce the recombinant protein production. Cells were collected 3 hours later by centrifugation. The bacterial pellet from 250 ml culture was suspended in 10 ml lysis buffer (Tris.HCl pH7.5, 1 mM DTT, 1 mM EDTA, 1 mM PMSF and 0.1% Triton X-100) and submitted to three freeze/thaw cycles before sonication. Cell lysate was clarified by centrifugation 20 minutes at 8000 rpm. The pellet collected was suspended again in extraction buffer, the resulting suspension sonicated, and pellet collected by centrifugation 20 minutes at 8000 rpm. A third wash was performed the same way with Tris extraction buffer + 1M NaCl and a fourth wash with Tris extraction buffer. After the different washing steps, the pellet contains *FL66* or *FL39* protein at 90% homogeneity. The pellets were suspended in Laemli loading buffer (Laemmli, 1970, Nature 277, p680-681) and *FL66* and *FL39* were further purified by SDS/ 12% polyacrylamide gel electrophoresis. The gel was stained in 0.025% coomassie brillant blue R250 in water and destained in water. The strong band co-migrating at the 31 kDa molecular weight marker position was cut out of the gel with a scalpel. The polyacrylamid fragments containing *FL66* or *FL39* were lyophilized and reduced into powder. The rabbit immunization was performed in complete Freund adjuvand, sub-cutaneous, with this antigen preparations. One injection corresponds to 100 µg of protein. The boosting injection were performed with non-complete Freund adjuvant, sub-cutaneous. The obtained sera detected bands of the expected size in protein extracts prepared from 2-day-old actively dividing cell cultures. No signals were observed using the pr-immune sera.

Example 6: Inhibition of kinase activity by FL66

The FL66pET/BL21(DE3) strain was used for the production of recombinant FL66. The inclusion bodies containing FL66 were collected and washed as described above. The recombinant FL66 protein was solubilized in 50 mM Tris.HCl pH7.6, 6M urea and kept on ice for 1 hour. Refolding of the FL66 protein was performed by removing urea on a sephadex G25 gel filtration column, equilibrated in 50 mM Tris.HCl pH7.6, 400 mM NaCl. The collected fractions were centrifuged and the supernatant was used for the inhibition assay.

CDK complexes from *A. thaliana* were purified on p13^{suc1} sepharose beads, starting from 100 µg of total protein extract prepared from a 2-day-old cell suspension culture. The FL66 protein was added to these purified complexes at a final concentration of 10 nM, 100 nM, 1 µM and 10 µM. After incubation during 1 hour on ice the CDK activity was measured using histone H1 as substrate, according to Azzi *et al.* (1992, Eur. J. Biochem., 203, 353-380). When compared to a control sample (without addition of FL66), the activity was found to be 82 % of the control after addition of 10 nM of FL66, 74 % after addition of 100 nM, 56 % after addition of 1 µM, and 12 % after addition of 10 µM of FL66. Addition of 30 µM of bovine serum albumin by comparison gives only a non-specific decrease to 70 % of the control activity.

The FL66 preparation was also added to *A. thaliana* CDK fraction bound to p13^{suc1} beads, prior to washing of these beads. The kinase activity dropped to 81 % and 35 % of the control with a concentration of 0.1 µM and 10 µM of FL66, respectively.

Example 7: The *Arabidopsis thaliana* CKI FL66 associates exclusively with CDC2aAt *in vitro*

Purified recombinant FL66 protein (prepared as described as in previous Example 6) was coupled to CNBr-activated Sepharose 4B (Pharmacia) at a concentration of 5 mg/ml of gel according to the manufacturer's instructions. Protein extracts were prepared from a 2-day-old cell suspension culture of *A. thaliana* Col-O in homogenisation buffer (HB) containing 50 mM Tris-HCl (pH 7.2), 60 mM β-

glycerophosphate, 15 mM nitrophenyl phosphate, 15 mM EGTA, 15 mM MgCl₂, 2 mM dithiothreitol, 0.1 mM vanadate, 50 mM NaF, 20 µg/ml leupeptin, 20 µg/ml aprotinin, 20 µg/ml soybean trypsin inhibitor (SBTI), 100 µM benzamidine, 1 mM phenylmethylsulfonylfluoride, and 0.1% Triton X-100. Two-hundred µg protein extract in a total volume of 100 µl HB was loaded on 50 µl 50% (v/v) FL66-Sepharose or control Sepharose beads, and incubated on a rotating wheel for 2 h at 4°C. The unbound proteins were collected for later analysis. The beads-bound fractions were washed 3 times with HB. Beads were resuspended in 30 µl SDS-loading buffer and boiled. The supernatants (beads bound fractions) and 10 µl of the unbound fractions were separated on a 12.5% SDS-PAGE gel and electroblotted on nitrocellulose membrane (Hybond-C⁺; Amersham). Filters were blocked overnight with 2% milk in phosphate buffered saline (PBS), washed 3 times with PBS, probed for 2 h with specific antibodies for CDC2aAt (1/5000 dilution) or CDC2bAt (1/2500 dilution) in PBS containing 0.5% Tween-20 and 1% albumin, washed for 1 h with PBS with 0.5% Tween-20, incubated for 2 h with peroxidase-conjugated secondary antibody (Amersham), and washed for 1 h with PBS containing 0.5% Tween 20. Protein detection was done by the chemoluminescent procedure (Pierce).

Western blotting revealed that the a significant fraction of CDC2aAt retained on the FL66-Sepharose beads, but not on the control beads, demonstrating the *in vitro* interaction between FL66 and CDC2aAt. In contrast, the CDC2bAt protein did not retain on the FL66-Sepharose beads but was found back in the unbound fraction. These results demonstrate the specificity of the FL66 protein for CDC2aAt.

Example 8: Expression of CKIs at different time-points in an asynchronous cell suspension culture of *Arabidopsis thaliana*

The expression levels of the different *A. thaliana* CKI genes (*FL39*, *FL66*, and *LDV159*) at different time-points during the cultivation of a *A. thaliana* cell culture were studied by reverse-transcriptase polymerase chain reaction (RT-PCR) technology. Four time-points were considered, representing the cell culture at different growth phases: day 1 (lag phase), day 5 (exponential growth phase), day 8 (beginning of the stationary phase), and day 12 (late stationary phase). Total RNA

of cells harvested at these time-points was extracted using the Trizol reagent (Gibco BRL). 75 µg of this total RNA preparation was used for mRNA extraction using Dynabeads oligoT25 (Dyna). This mRNA was used to prepare cDNA using the universal riboclone cDNA synthesis system (Promega). Five ng of cDNA was subsequently used for RT-PCR, using 300 ng of each of the appropriate forward and reverse primers, 160 µM of dNTPs, 10 µl of PCR buffer, and 0.8 µl of Taq polymerase (Promega). The used primers were 5'-CGGCTCGAGGAGAACCACAAACACGC-3' (SEQ ID NO: 13) and 5'-CGAAACTAGTTAATTACCTCAAGGAAG-3' (SEQ ID NO: 14) for *FL39*; 5'-GATCCCGGGCGATATCAGCGTCATGG-3' (SEQ ID NO: 15) and 5'-GATCCCGGGTTAGTCTGTAACTCC-3' (SEQ ID NO: 16) for *FL66*; 5'-GCAGCTACGGAGCCGGAGAATTGT-3' (SEQ ID NO: 17) and 5'-TCTCCTTCTCGAAATCGAAATTGTACT-3' (SEQ ID NO: 18) and for *LDV159*. The PCR reaction consisted of 4 min preheating at 94°C, followed by cycles of 45 sec 94°C, 45 sec 45°C, and 45 sec 72°C. After 10, 15, 20, 25, 30 and 35 cycles 10 µl of the amplification mixture was loaded on an agarose gel and electrophoretically separated. After depurination, denaturation, and neutralisation of the DNA it was transferred to a nitro-cellulose membrane (Hybond N⁺; Amersham). The DNA was fixed on the membrane by UV crosslinking.

Membranes were hybridised using fluorescein-labelled probes prepared of the *FL39*, *FL66*, or *LDV159* genes according to the manufacturer's protocol (Amersham). After 16 hours hybridisation at 65°C, the membranes were washed for 15 min using 2X SSC; 0.1X SDS, and 15 min using 1X SSC; 0.1X SDS. The signals were detected using the CDP-star detection module according to the manufacturer's protocol (Amersham). The signals were revealed by autoradiography.

FL39 transcripts could be detected at days 1, 5, and 8; but not in late stationary cells (day 12). The strongest expression was noticed in cells being in the exponential growth phase (at day 5). The *FL66* and *LDV159* genes were most abundantly expressed at day 5 (during the exponential growth phase), although expression was already substantial high at day 1 during the lag phase. Both genes were expressed at a strongly reduced level in stationary cultures (at day 8 and 12).

Example 9: *FL66* transcription is upregulated by NaCl

Stationary *A. thaliana* suspension cultures were diluted at day 1 in fresh medium and cultivated for 48 hours. At this time-point the culture was divided into two subcultures. At one of these cultures 1% NaCl was added. The cultures were cultivated for 12 hours after which the cells were collected and frozen in liquid nitrogen. Of these samples RNA was prepared using the Trizol reagent (Bibco BRL). 100 µg of this total RNA preparation of both samples was used for mRNA extraction using Dynabeads oligoT25 (Dyna). The poly-A RNA was electrophorically separated on an agarose gel and blotted onto a nitro-cellulose membrane (Hybond-N⁺, Amersham). The membrane was hybridised using a fluorescein-labelled probe prepared of the *FL66* sequence according to the manufacturer's protocol (Amersham). After 16 hours hybridisation at 65°C, the membranes were washed for 15 min using 2X SSC; 0.1X SDS, and 15 min using 1X SSC; 0.1X SDS. The signals were detected using the CDP-star detection module (Amersham). The signals were revealed by autoradiography.

A weak hybridising band of approximately 1000 bp was detected in the control sample. Treatment with 1% NaCl clearly increased the intensity of the hybridisation signal. This demonstrates that the stress caused by the addition of NaCl results in the transcriptional activation of the *FL66* gene. This induction could result in a permanent or transient arrest of cell division activity.

Example 10: Production of the CKIs in plants

To obtain transgenic plants overexpressing the *A. thaliana* CKI genes, the coding regions of *FL36*, *FL66*, and *LDV159* were cloned into the pAT7002 vector (Aoyama and Chua, 1997, Plant J. 11, p605-612). This vector allows inducible expression of the cloned inserts by the addition of the glucocorticoid dexamethasone. Following the polymerase chain reaction (PCR) technology the coding regions of *FL39*, *FL66*, and *ICK1* were amplified using the appropriate primer combinations. The primers used were 5'-CGGCTCGAGGAGAACCACAAACACGC-3' (SEQ ID NO: 19) and 5'-CGAAACTAGTTAATTACCTCAAGGAAG-3' (SEQ ID NO: 20) for *FL39*,

GATCCCGGGCGATATCAGCGTCATGG-3' (SEQ ID NO: 21) and 5'-GATCCCGGGTTAGTCTGTAACTCC-3' (SEQ ID NO: 22) for *FL66*, and 5'-CCCGCTCGAGATGGTGAGAAAATATAGAAAAGCTAAAGGATTTGTAGAAGCTG GAGTTTCGTCAACGTA-3' (SEQ ID NO: 23) and 5'-GGACTAGTTCACTCTAACTTTACCCATTCG-3' (SEQ ID NO: 24) for *LDV159*. The obtained *FL39* and *LDV159* PCR fragments were purified and cut with *XhoI* and *SpeI*. Subsequently these fragments were used to clone into the *XhoI* and *SpeI* sites of pTA7002. The obtained *FL66* fragment was cut with *SmaI*, purified, and cloned blunt into the *XhoI* and *SpeI* sites of the pTA7002 vector. The resulted binary vectors were transferred into *Agrobacterium tumefaciens*. These strains were used to transform *Nicotiana tabacum* cv. Petit havana using the leaf disk protocol (Horsh *et al.*, 1985, Science 227, p1229-1231) and *Arabidopsis thaliana* using the root transformation protocol (Valvekens *et al.*, 1988, PNAS 85, p5536-5540).

Example 11: *Arabidopsis thaliana* CKIs expression in fission yeast *Schizosaccharomyces pombe*

To obtain heterologous expression of *A. thaliana* CKI genes in the fission yeast *Schizosaccharomyces pombe*, the *FL39* and *FL66* were cloned into the pREP81 (Basi *et al.*, 1993, Gene 123, p131-136) and BNRP3 (Hemerly *et al.*, 1995, EMBO J. 14, p3925-3936) vectors. These vectors contain the thiamine-repressible promoter *mtl1* and allow inducible expression of the *FL39* and *FL66* genes (Maundrell *et al.*, 1990, JBC 265, p10857-10864). The expression is inducible to different levels: strong induction is obtained with BNRP3, low induction with pREP81. The coding region of *FL39* and *FL66* were amplified by PCR technique. The *FL66* amplified coding sequence encloses a protein starting at the methionine at amino-acid position 11, therefore not including the first 10 amino-acids encoded by the *FL66* clone. Primers used to amplify *FL39* were 5'-GATCATCTTAAGCATCATCGTCTTCTTCATGG-3' (SEQ ID NO: 25) and 5'-TAGGAGCATATGGCGGCGG-3' (SEQ ID NO: 26). Primers used to amplify *FL66* were 5'-ATATCAGCGCCATGGAAGTC-3' (SEQ ID NO: 27) and 5'-GGAGCTGGATCCTTTTGAATTCATGG-3' (SEQ ID NO: 28). The obtained *FL39*

PCR fragment was purified, phosphorylated with polynucleotide kinase (blunt end) and cut with *NdeI*. This fragment was cloned into the *NdeI* and *SmaI* sites of pREP81. The obtained *FL66* PCR fragment was purified, cut with *NcoI* and *BamHI* and cloned into the *NcoI* and *BamHI* sites of BNRP3.

The resulting recombinant plasmids were transformed in 972 *leu1-32 h⁻* *Sch. pombe* strain (wild type) by electroporation technique. Transformant were selected on inducing medium supplemented with 5 µg/ml of thiamine. Phenotypes of transformants were then compared with the phenotype of wild type strain, on non-inducing medium. No cell cycle block could be observed in *Sch. pombe* transformants expressing *FL39* or *FL66*.

Example 12: Identification of a *FL66* related gene

By screening the *A. thaliana* sequence databank a genomic sequence was identified encoding a protein highly homologous to *FL66*. The protein encoded, annotated as 'unknown protein', was renamed *FL67*. *FL67* shows 39.545 % similarity and 30.909 % identity with *FL66*.

Example 13: In situ hybridisation patterns

Plant material was fixed in 2.5% glutaraldehyde in 0.1M cacodylate buffer (pH7.2) and dehydrated until 100% ethanol prior to embedding in paraffin and tissue sectioning. ³⁵S-UTP-labeled sense and antisense RNAs of cDNA from *FL39*, *FL66* and *LDV159* subcloned in PGem2 were generated by run-off transcription using T7 and Sp6 RNA polymerases according to the manufacturer's instructions (Boehringer Mannheim). Labeled RNA probes were hydrolysed to an average length of 200 nt according to Cox et al (1984). Deparaffinized and rehydrated tissue sections were taken through the mRNA in situ procedure essentially as described by Angerer and Angerer (1992). Stringencies during washes were 2X SSC at room temperature for 60 min and 0.1X SSC in 50% formamide at 45 C for 30 min. RNase treatment, washing steps, photograph emulsion coating, and the development of slides were performed as described by Angerer and Angerer (1992). Photographs were taken

with a Diaplan microscope equipped with dark-field optics (Leitz, Wetzlar, Germany).

Distinct expression patterns of the *FL39*, *LDV159* and *FL66* genes were observed when applying the mRNA in situ hybridization technique on *Arabidopsis thaliana* and radish seedlings. Sections of paraffin embedded roots, shoot apical meristems, flowers and siliques of *Arabidopsis thaliana*, and radish roots and shoot apical meristems were used to hybridize with the three cyclin-dependent kinase inhibitors. The *FL39* gene is expressed in young root meristems in a homogeneous pattern. Mature root meristems barely showed any expression of the gene. Some regions along the root vascular tissue showed alternating zones of expressing and nonexpressing cells at the periphery of the vascular bundle. A region of pericycle cells in the vascular tissue, flanking the region where new lateral roots usually form, presented a very strong expression of the *FL39* gene. In contrast, pericycle cells on the region where lateral roots form hardly showed any expression. These results show that higher levels of *FL39* mRNA was observed close to the region where lateral roots emerge possibly preventing their formation at these regions. On the other hand, the absence of *FL39* gene expression in the poles of the diarch vascular bundle may allow lateral root formation at these sites. It possibly assures that lateral roots are formed by division of pericycle cells adjacent to a protoxylem group. Uniform expression of *FL39* gene was also observed in all cells of the shoot apical meristem. Strong signals were observed at the surface and tip of young leaves. The epidermal and palisade layers of the leaves are the first layers to vacuolize and differentiate, and the oldest part of the leaves are at the tip. In addition, the expression pattern of *CYCB1;1*, a molecular marker of cell division, shows a basipetal pattern of cessation of cell division. Therefore, *FL39* expression at these sites may inhibit cell division allowing cell differentiation to occur during early stages of leaf development. A similar pattern of expression was observed on radish leaves, roots and shoot apical meristems. In addition, strong expression at the epidermis of the stem was also observed on young seedlings. The presence of *FL39* mRNA in these cells might allow cells to differentiate. In *Arabidopsis* flowers, *FL39* was mainly expressed in the tapetal layer of the anthers and in pollen grains.

Considering that at this stage, tapetum and pollen grains do not divide, *FL39* might be expressed at these sites to inhibit cell division. Weaker expression was observed in flower buds and mature ovaries. During embryo development very strong expression was observed in embryos at the globular, heart and torpedo stages. At the later stage strongest expression was at the embryonic root. Weak or no hybridization signal was observed in mature seeds.

Expression of the *LDV159* was also observed in all cells along the main and lateral root meristems and shoot apical meristems, but in a more uniform manner. Expression in vascular tissue was slightly patchy, and stronger at the pericycle. Often a patchy pattern was observed in distinct cells of mature leaves. In flowers, expression was mainly observed in mature ovaries. Expression in embryos was mainly observed in globular and heart stages and in the embryonic root at the torpedo stage. Weak expression was observed in mature embryos. These results suggest a function of *LDV159* in the regulation of correct progression through the cell cycle. *LDV159* might play a role in the checkpoint control, avoiding the premature activation of the CDK complexes under unfavorable conditions. Its association with CDKs could inhibit CDK activity until the cell perceives the correct signals to progress to the next cell cycle phase.

FL66 gene expression was observed in the root and shoot apical meristems. Stronger expression was observed in young differentiating leaves often in a patchy manner suggesting a cell cycle phase dependent expression pattern. Hybridization signal was also observed along the vascular tissue. *FL66* expression was as well observed in flower buds and young flowers. In mature flowers stronger expression was observed in the ovary wall, funiculus, ovules and pollen grains. During embryo development strong expression was observed at the globular stage. Signal gradually decreases until the embryo maturation. Stronger signals were often observed in the embryonic root.

Example 14: Identification of a CKI in alfalfa

The *Medicago sativa* cdc2-related kinase (CDC2AMs; Magyar et al., 1997., The Plant Cell, Vol.: 9, 223-235.) cloned in the vector pBD-GAL4 Cam phagemid (Stratagene) was used as a bait protein in a yeast two-hybrid screen. mRNA isolated from young alfalfa (*Medicago truncatula*) root nodules was converted to cDNA followed by cloning into HybridZAP phagemids (Stratagene). The library was converted to pAD-GAL4 plasmid library by mass excision. The yeast strain Y190 (Clontech) was used as a host for the two hybrid analysis. As a positive clone interacting in this system with the CDC2MsA kinase, a partial cDNA clone of 613bp was isolated coding for 128 amino acids. Sequencing of this clone revealed extensive homology with the C-terminal region of known CDK inhibitors (CKI). The full length cDNA clone was isolated with screening an alfalfa root nodule Lambda ZAP II (Stratagene) cDNA library with the partial cDNA as probe and using standard procedures. A clone comprising a full length cDNA designated ALFCDKI was obtained and the corresponding nucleotide and amino acid sequences of the encoded CKI are shown in SEQ ID NO: 5 and 6, respectively.

61
SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Cropdesign NV
(B) STREET: Technologiepark 3
(C) CITY: Zwijnaarde
(D) STATE: none
(E) COUNTRY: Belgium
(F) POSTAL CODE (ZIP): 9052

(ii) TITLE OF INVENTION: Cyclin-dependent kinase inhibitors and uses thereof

(iii) NUMBER OF SEQUENCES: 33

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 86..712

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCACGAGGA GAACCACAAA CACGCACACA TAACGAGTGA TTTTAGAGAG AGATAGAGAT	60
CTGGAAGGTG ACGTCGTAGG AGATT ATG GCG GCG GTT AGG AGA AGA GAA CGA	112
Met Ala Ala Val Arg Arg Arg Glu Arg	
1 5	
GAT GTG GTT GAA GAG AAT GGA GTT ACG ACG ACG ACG GTG AAA CGA AGG	160
Asp Val Val Glu Glu Asn Gly Val Thr Thr Thr Val Lys Arg Arg	
10 15 20 25	
AAG ATG GAG GAG GAA GTG GAT TTA GTG GAA TCT AGG ATA ATT CTG TCT	208
Lys Met Glu Glu Glu Val Asp Leu Val Glu Ser Arg Ile Ile Leu Ser	
30 35 40	

CCG TGT GTA CAG GCG ACG AAT CGC GGT GGA ATT GTG GCG AGA AAT TCA	256
Pro Cys Val Gln Ala Thr Asn Arg Gly Gly Ile Val Ala Arg Asn Ser	
45 50 55	
GCA GGA GCG TCG GAG ACG AGT GTT GTT ATA GTA CGA CGG CGA GAT TCT	304
Ala Gly Ala Ser Glu Thr Ser Val Val Ile Val Arg Arg Arg Asp Ser	
60 65 70	
CCT CCG GTT GAA GAA CAG TGT CAA ATC GAA GAA GAA GAT TCG TCG GTT	352
Pro Pro Val Glu Glu Gln Cys Gln Ile Glu Glu Glu Asp Ser Ser Val	
75 80 85	
TCG TGT TGT TCT ACA TCG GAA GAG AAA TCG AAA CGG AGA ATC GAA TTT	400
Ser Cys Cys Ser Thr Ser Glu Glu Lys Ser Lys Arg Arg Ile Glu Phe	
90 95 100 105	
GTA GAT CTT GAG GAA AAT AAC GGT GAC GAT CGT GAA ACA GAA ACG TCG	448
Val Asp Leu Glu Glu Asn Asn Gly Asp Asp Arg Glu Thr Glu Thr Ser	
110 115 120	
TGG ATT TAC GAT GAT TTG AAT AAG AGT GAG GAA TCG ATG AAC ATG GAT	496
Trp Ile Tyr Asp Asp Leu Asn Lys Ser Glu Glu Ser Met Asn Met Asp	
125 130 135	
TCT TCT TCG GTG GCT GTT GAA GAT GTA GAG TCT CGC CGC AGG TTA AGG	544
Ser Ser Ser Val Ala Val Glu Asp Val Glu Ser Arg Arg Arg Leu Arg	
140 145 150	
AAG AGT CTC CAT GAG ACG GTG AAG GAA GCT GAG TTA GAA GAT TTT TTT	592
Lys Ser Leu His Glu Thr Val Lys Glu Ala Glu Leu Glu Asp Phe Phe	
155 160 165	
CAG GTG GCG GAG AAA GAT CTT CGG AAT AAG TTG TTG GAA TGT TCT ATG	640
Gln Val Ala Glu Lys Asp Leu Arg Asn Lys Leu Leu Glu Cys Ser Met	
170 175 180 185	
AAG TAT AAC TTC GAT TTC GAG AAA GAT GAG CCA CTT GGT GGA GGA AGA	688
Lys Tyr Asn Phe Asp Phe Glu Lys Asp Glu Pro Leu Gly Gly Gly Arg	
190 195 200	
TAC GAG TGG GTT AAA TTG AAT CCA TGAAGAAGAC GATGATGATA ATGATGATCA	742
Tyr Glu Trp Val Lys Leu Asn Pro	
205	
TTGTTTTTAC CAAAGTACTT ATTATTTTTC TTCTGTAATA ATCTTTGCTT TGATTTTTCT	802
TTTAACAAAA TCCAAATGTA GATATCTTTC TCTCGAATAA TCAATAACAT GTAATTCAAC	862
TTTTGTTTGT ACTTCCTTGA GGTAATTAAT TAGATTCGTG TTTTCTCGA TTAATAAACT	922
ATAAGTTTAT	932

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 209 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Ala Ala Val Arg Arg Arg Glu Arg Asp Val Val Glu Glu Asn Gly
 1             5             10             15

Val Thr Thr Thr Thr Val Lys Arg Arg Lys Met Glu Glu Glu Val Asp
      20             25             30

Leu Val Glu Ser Arg Ile Ile Leu Ser Pro Cys Val Gln Ala Thr Asn
      35             40             45

Arg Gly Gly Ile Val Ala Arg Asn Ser Ala Gly Ala Ser Glu Thr Ser
      50             55             60

Val Val Ile Val Arg Arg Arg Asp Ser Pro Pro Val Glu Glu Gln Cys
      65             70             75             80

Gln Ile Glu Glu Glu Asp Ser Ser Val Ser Cys Cys Ser Thr Ser Glu
      85             90             95

Glu Lys Ser Lys Arg Arg Ile Glu Phe Val Asp Leu Glu Glu Asn Asn
      100            105            110

Gly Asp Asp Arg Glu Thr Glu Thr Ser Trp Ile Tyr Asp Asp Leu Asn
      115            120            125

Lys Ser Glu Glu Ser Met Asn Met Asp Ser Ser Ser Val Ala Val Glu
      130            135            140

Asp Val Glu Ser Arg Arg Arg Leu Arg Lys Ser Leu His Glu Thr Val
      145            150            155            160

Lys Glu Ala Glu Leu Glu Asp Phe Phe Gln Val Ala Glu Lys Asp Leu
      165            170            175

Arg Asn Lys Leu Leu Glu Cys Ser Met Lys Tyr Asn Phe Asp Phe Glu
      180            185            190

Lys Asp Glu Pro Leu Gly Gly Gly Arg Tyr Glu Trp Val Lys Leu Asn
      195            200            205

```

Pro

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 875 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:11..658

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCACGAGAG AAA TCA AAG ATA ACT GGC GAT ATC AGC GTC ATG GAA GTC	49
Lys Ser Lys Ile Thr Gly Asp Ile Ser Val Met Glu Val	
210 215 220	
TCT AAA GCA ACA GCT CCA AGT CCA GGT GTT CGA ACC AGA GCC GCT AAA	97
Ser Lys Ala Thr Ala Pro Ser Pro Gly Val Arg Thr Arg Ala Ala Lys	
225 230 235	
ACC CTA GCC TTG AAG CGG CTT AAT TCC TCC GCC GCT GAT TCA GCT CTA	145
Thr Leu Ala Leu Lys Arg Leu Asn Ser Ser Ala Ala Asp Ser Ala Leu	
240 245 250	
CCT AAC GAC TCT TCT TGC TAT CTT CAG CTC CGT AGC CGC CGT CTC GAG	193
Pro Asn Asp Ser Ser Cys Tyr Leu Gln Leu Arg Ser Arg Arg Leu Glu	
255 260 265 270	
AAA CCC TCT TCG CTG ATT GAA CCG AAA CAG CCG CCG AGA GTT CAC AGA	241
Lys Pro Ser Ser Leu Ile Glu Pro Lys Gln Pro Pro Arg Val His Arg	
275 280 285	
TCG GGA ATT AAA GAG TCT GGT TCC AGG TCT CGC GTT GAC TCG GTT AAC	289
Ser Gly Ile Lys Glu Ser Gly Ser Arg Ser Arg Val Asp Ser Val Asn	
290 295 300	
TCG GTT CCT GTA GCT CAG AGC TCT AAT GAA GAT GAA TGT TTT GAC AAT	337
Ser Val Pro Val Ala Gln Ser Ser Asn Glu Asp Glu Cys Phe Asp Asn	
305 310 315	
TTC GTG AGT GTC CAA GTT TCT TGT GGT GAA AAC AGT CTC GGT TTT GAA	385
Phe Val Ser Val Gln Val Ser Cys Gly Glu Asn Ser Leu Gly Phe Glu	
320 325 330	
TCA AGA CAC AGC ACA AGG GAG AGC ACG CCT TGT AAC TTT GTT GAG GAT	433
Ser Arg His Ser Thr Arg Glu Ser Thr Pro Cys Asn Phe Val Glu Asp	
335 340 345 350	
ATG GAG ATC ATG GTT ACA CCA GGG TCT AGC ACG AGG TCG ATG TGC AGA	481
Met Glu Ile Met Val Thr Pro Gly Ser Ser Thr Arg Ser Met Cys Arg	
355 360 365	
GCA ACC AAA GAG TAC ACA AGG GAA CAA GAT AAC GTG ATC CCG ACC ACT	529
Ala Thr Lys Glu Tyr Thr Arg Glu Gln Asp Asn Val Ile Pro Thr Thr	
370 375 380	

AGT GAA ATG GAG GAG TTC TTT GCA TAT GCA GAG CAG CAG CAA CAG AGG	577
Ser Glu Met Glu Glu Phe Phe Ala Tyr Ala Glu Gln Gln Gln Gln Arg	
385 390 395	
CTA TTC ATG GAG AAG TAC AAC TTC GAC ATT GTG AAT GAT ATC CCC CTC	625
Leu Phe Met Glu Lys Tyr Asn Phe Asp Ile Val Asn Asp Ile Pro Leu	
400 405 410	
AGC GGA CGT TAC GAA TGG GTG CAA GTC AAA CCA TGAAGTTCAA AAGGAAACAG	678
Ser Gly Arg Tyr Glu Trp Val Gln Val Lys Pro	
415 420 425	
CTCCAAAAGA CATGGTGTGA AGTTAGAGAA TGTGATGGAG TTAACAGACT AACCAAACAT	738
CAGAAATCGT GTAATCTTAA GTAATAATGT GGTAGAGAA CAAGTTTGAG AGTAGCTTAG	798
GGACCTTAAA ACCTCACACC ATTTGTAATA CTAATCTTCT TCAGATGCTT AGTGAAATTT	858
TCTCATCTGT TTCTTTC	875

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 216 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Lys Ser Lys Ile Thr Gly Asp Ile Ser Val Met Glu Val Ser Lys Ala	
1 5 10 15	
Thr Ala Pro Ser Pro Gly Val Arg Thr Arg Ala Ala Lys Thr Leu Ala	
20 25 30	
Leu Lys Arg Leu Asn Ser Ser Ala Ala Asp Ser Ala Leu Pro Asn Asp	
35 40 45	
Ser Ser Cys Tyr Leu Gln Leu Arg Ser Arg Arg Leu Glu Lys Pro Ser	
50 55 60	
Ser Leu Ile Glu Pro Lys Gln Pro Pro Arg Val His Arg Ser Gly Ile	
65 70 75 80	
Lys Glu Ser Gly Ser Arg Ser Arg Val Asp Ser Val Asn Ser Val Pro	
85 90 95	
Val Ala Gln Ser Ser Asn Glu Asp Glu Cys Phe Asp Asn Phe Val Ser	
100 105 110	
Val Gln Val Ser Cys Gly Glu Asn Ser Leu Gly Phe Glu Ser Arg His	
115 120 125	
Ser Thr Arg Glu Ser Thr Pro Cys Asn Phe Val Glu Asp Met Glu Ile	
130 135 140	

Met Val Thr Pro Gly Ser Ser Thr Arg Ser Met Cys Arg Ala Thr Lys
 145 150 155 160

Glu Tyr Thr Arg Glu Gln Asp Asn Val Ile Pro Thr Thr Ser Glu Met
 165 170 175

Glu Glu Phe Phe Ala Tyr Ala Glu Gln Gln Gln Gln Arg Leu Phe Met
 180 185 190

Glu Lys Tyr Asn Phe Asp Ile Val Asn Asp Ile Pro Leu Ser Gly Arg
 195 200 205

Tyr Glu Trp Val Gln Val Lys Pro
 210 215

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 92..763

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AAACCACTCT TCAAATCAAA CACTTTCTTA CATAAGATTC CTCTGTTTTT CTGTGTGCTT 60

CTTCAAATTC TTCCCTGTT TTTCAACTTC A ATG GGG AAG TAC ATG AAG AAA 112
 Met Gly Lys Tyr Met Lys Lys
 220

CTC AAA TCC AAA TCA GAA TCT CCT TCA CCC AAT TCA ACA CCA ACA CCA 160
 Leu Lys Ser Lys Ser Glu Ser Pro Ser Pro Asn Ser Thr Pro Thr Pro
 225 230 235

TCA CCA TCA CCA TCA CCA ACA CCA ATC ACC ACC AAT TCA CCA CCA CCA 208
 Ser Pro Ser Pro Ser Pro Thr Pro Ile Thr Thr Asn Ser Pro Pro Pro
 240 245 250 255

ACA ACA CCC AAT TCC TCT GAT GGT GTT CGA ACT CGT GCT AGA ACC CTA 256
 Thr Thr Pro Asn Ser Ser Asp Gly Val Arg Thr Arg Ala Arg Thr Leu
 260 265 270

GCT TTG GAG AAT TCC AAC AAT CAG AAT CAG AAT CTT TCT GTT TCT TCT 304
 Ala Leu Glu Asn Ser Asn Asn Gln Asn Gln Asn Leu Ser Val Ser Ser
 275 280 285

GAT TCT TAC CTT CAG CTG AGG AAC CGT CGC CTT AAG AGA CCC CTA ATT Asp Ser Tyr Leu Gln Leu Arg Asn Arg Arg Leu Lys Arg Pro Leu Ile 290 295 300	352
AGG CAA CAT TCC GCT AAG AGG AAT AAG GGG CAT GAT GGA AAC CCT AAA Arg Gln His Ser Ala Lys Arg Asn Lys Gly His Asp Gly Asn Pro Lys 305 310 315	400
TCC CCA ATT GGG GAT TCA ATT GCT GAA GAG AAA ACT GTT CAG AAG AGT Ser Pro Ile Gly Asp Ser Ile Ala Glu Glu Lys Thr Val Gln Lys Ser 320 325 330 335	448
CCT GAG CCT GAA AAT GCT GAA TTC AAG GAG AAT GCT GAG GAT ACT GAG Pro Glu Pro Glu Asn Ala Glu Phe Lys Glu Asn Ala Glu Asp Thr Glu 340 345 350	496
AGA AGC GCT AGG GAA ACT ACA CCC GTC CAT TTG ATA ATG CGA GCA GAC Arg Ser Ala Arg Glu Thr Thr Pro Val His Leu Ile Met Arg Ala Asp 355 360 365	544
GTT CTC AGG CCT CCT AGG CCA ATT ACC AGG CGT ACT TTT CCA ACT GAA Val Leu Arg Pro Pro Arg Pro Ile Thr Arg Arg Thr Phe Pro Thr Glu 370 375 380	592
GCT AAT CCC AAA ACG GAG CAG CCA ACT ATC CCA ATT TCA CGC GAA TTT Ala Asn Pro Lys Thr Glu Gln Pro Thr Ile Pro Ile Ser Arg Glu Phe 385 390 395	640
GAG GAA TTC TGT GCT AAA CAT GAA GCC GAG CAG CAA AGG GAG TTC ATG Glu Glu Phe Cys Ala Lys His Glu Ala Glu Gln Gln Arg Glu Phe Met 400 405 410 415	688
GAG AAG TAC AAC TTT GAT CCT GTG ACA GAG CAG CCA CTC CCA GGG CGT Glu Lys Tyr Asn Phe Asp Pro Val Thr Glu Gln Pro Leu Pro Gly Arg 420 425 430	736
TAC GAA TGG GAA AAA GTG TCG CCC TAG AAGGCAGGCT AGTATTAAGT Tyr Glu Trp Glu Lys Val Ser Pro *	783
435 440	
GTTCCATCAA TACATCTTTA AAGTAGCAGC AGGGTTAGAA TTTGTTGAAA AGGGTGGTGG	843
TGCTATTTCC ATTTTCCATC ACTTTCTATT TACTTGTAAG GAAAGTAGGA CTTTCAACAT	903
ATGTAGACTA ATGATCTGTA ACTTTACAGA GGTGTTGATT ACACAACAAT ACAAAGTCCT	963
TTGTCTAGCA GATCATTAAA GAAGGGTTTG AGGGAATAAG GGTCTCTAGT TGTAGGGTTT	1023
AGGGTATAAA ATCAAAGTAG GGTATGTAAG AGAGGTTTTA CAAGAATTTC CTTTTGTTCT	1083
TGTGTTTTAC TCTTGTTTTG TCTATACTTG TACTCATGGA ACTTCAACAA ACTCTTAAGA	1143
AATAAAGAAC CAGATCTCCC TCAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1193

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 224 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```

Met Gly Lys Tyr Met Lys Lys Leu Lys Ser Lys Ser Glu Ser Pro Ser
 1           5           10           15

Pro Asn Ser Thr Pro Thr Pro Ser Pro Ser Pro Ser Pro Thr Pro Ile
          20           25           30

Thr Thr Asn Ser Pro Pro Pro Thr Thr Pro Asn Ser Ser Asp Gly Val
          35           40           45

Arg Thr Arg Ala Arg Thr Leu Ala Leu Glu Asn Ser Asn Asn Gln Asn
 50           55           60

Gln Asn Leu Ser Val Ser Ser Asp Ser Tyr Leu Gln Leu Arg Asn Arg
 65           70           75           80

Arg Leu Lys Arg Pro Leu Ile Arg Gln His Ser Ala Lys Arg Asn Lys
          85           90           95

Gly His Asp Gly Asn Pro Lys Ser Pro Ile Gly Asp Ser Ile Ala Glu
          100          105          110

Glu Lys Thr Val Gln Lys Ser Pro Glu Pro Glu Asn Ala Glu Phe Lys
          115          120          125

Glu Asn Ala Glu Asp Thr Glu Arg Ser Ala Arg Glu Thr Thr Pro Val
          130          135          140

His Leu Ile Met Arg Ala Asp Val Leu Arg Pro Pro Arg Pro Ile Thr
          145          150          155          160

Arg Arg Thr Phe Pro Thr Glu Ala Asn Pro Lys Thr Glu Gln Pro Thr
          165          170          175

Ile Pro Ile Ser Arg Glu Phe Glu Glu Phe Cys Ala Lys His Glu Ala
          180          185          190

Glu Gln Gln Arg Glu Phe Met Glu Lys Tyr Asn Phe Asp Pro Val Thr
          195          200          205

Glu Gln Pro Leu Pro Gly Arg Tyr Glu Trp Glu Lys Val Ser Pro *
          210          215          220

```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGAGATCTGA ATTCATGGAT CAGTA

25

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGAGATCTGA ATTCCTAAGG CATGCC

26

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGGAATCCAT GGGCGGCGGT TAGGAGAAG

29

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGCGGATCCC GTCTTCTTCA TGGATTC

27

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGCGAATCCA TGGAAGTCTC TAAAGCAAC

29

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGCGGATCCT TTTGAACTTC ATGGTTTGAC

30

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CGGCTCGAGG AGAACCACAA ACACGC

26

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CGAAACTAGT TAATTACCTC AAGGAAG

27

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GATCCCGGGC GATATCAGCG TCATGG

26

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GATCCCGGGT TAGTCTGTTA ACTCC

25

- (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCAGCTACGG AGCCGGAGAA TTGT

24

- (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GCAGCTACGG AGCCGGAGAA TTGT

24

- (2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCTCCTTCTC GAAATCGAAA TTGTACT

27

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CGGCTCGAGG AGAACCACAA ACACGC

26

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"

- (iii) HYPOTHETICAL: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CGAAACTAGT TAATTACCTC AAGGAAG

27

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"
- (iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GATCCCGGGC GATATCAGCG TCATGG

26

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"
- (iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GATCCCGGGT TAGTCTGTTA ACTCC

25

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCCCTCGAG ATGGTGAGAA AATATAGAAA AGCTAAAGGA TTTGTAGAAG CTGGAGTTTC

60

GTCAACGTA

69

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGACTAGTTC ACTCTAACTT TACCCATTCG

30

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TAGGAGCATA TGGCGGCGG

19

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATATCAGCGC CATGGAAGTC

20

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGAGCTGGAT CCTTTTGGA TTCATGG

27

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TAGGAGCATA TGGCGGCGG

19

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

ATCATCGAAT TCTTCATGGA TTC

23

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

ATATCAGCGC CATGGAAGTC

20

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGAGCTGGAT CCTTTTGGAA TTCATGG

27

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

VRRRD/EXXXVEE

11